CONTRIBUTIONS
IN
STATISTICS AND AGRICULTURAL SCIENCES

Presented to

DR V. G. PANSE

ON HIS 62nd BIRTHDAY

Edited by

G. R. SETH

In Collaboration with

J. S. SARMA
V. N. AMBLE
D. SINGH
M. N. DAS

INDIAN SOCIETY OF AGRICULTURAL STATISTICS
NEW DELHI

V. G. Panse

1968

Designed and Printed by
R. K. **Printers,** 80-D Kamla Nagar, Delhi-7

C. SUBRAMANIAM, B.A. BL.

PHONE : 440556
'RIVERVIEW'
GUINDY, MADRSAS-25

2nd Jan 1968

Dear Shri Seth,

I am glad to learn that the Council of the Indian Society of Agricultural Statistics is honouring Dr. V. G. Panse on the occasion of his sixty second birth day on 10th of Jan 1968. Shri Panse had done great service in promoting the use of modern statistical methods in the field of Agricultural Research. His was a pioneering work and has set a new path for the coming generation to follow

I wish the function every success Owing to other preoccupation I am unable to be present on the occasion

Yours sincerely

C Subramaniam

MINISTER OF TRANSPORT & SHIPPING
INDIA
Prof. V. K. R. V. Rao
NEW DELHI-1

January 5, 1968

MESSAGE

I deem it a privilege to pay my tribute of esteem and affection to Dr. V. G. Panse on the occasion of the celebration of his 62nd birthday. Dr. Panse is a distinguished and fearless scientist who has made a notable contribution to the agricultural world by his painstaking, objective and thorough studies on various botanical, statistical and economic problems relating to agriculture. In addition, he is a fine human being.

May God give him many more years of health, life, and service to the discipline to which he has dedicated himself.

(V. K. R. V. Rao)

MINISTER OF STATE FOR FOOD & AGRICULTURE
GOVERNMENT OF INDIA

New Delhi, the 3rd January, 1968

MESSAGE

I am very happy to know that the Indian Society of Agricultural Statistics is honouring Dr. V. G. Panse on his sixty-second birthday by presenting him a volume of contributed papers by Statisticians, Agricultural Economists and Agricultural Scientists. This indeed is a befitting tribute to the person who has done so much for the Society and for the development of agricultural statistics needed for research and planning. His contributions have earned him world-wide recognition. I hope that Dr. Panse will soon resume his devoted service to the field of agricultural statistics which has been interrupted by his recent illness. I wish Dr. Panse many more years of sound health and creative activity.

Annasaheb P. Shinde

(Annasaheb P. Shinde)

B. SIVARAMAN
सचिव
कृषि और सामुदायिक विकास
खाद्य, कृषि, सामुदायिक विकास तथा सहकारिता मंत्रालय
भारत सरकार
SECRETARY
AGRICULTURE & COMMUNITY DEVELOPMENT
MINISTRY OF FOOD, AGRICULTURE, COMMUNITY
DEVELOPMENT & CO-OPERATION
GOVERNMENT OF INDIA

NEW DELHI
1st January, 1968

MESSAGE

I very much welcome the initiative taken by the Indian Society of Agricultural Statistics in bringing out a volume of contributed articles on statistics, agricultural economics and agricultural sciences to be presented to Dr. V. G. Panse on his sixty-second birthday. In addition to contributing significantly to statistical research in relation to agricultural and biological sciences, Dr. Panse has rendered yeoman service to the cause of improvement of agricultural statistics. In recent years he has also taken increased interest in studies on planning for agricultural development. He has always brought to bear his extraordinary commonsense, critical approach and dynamic attitude on any work with which he was associated. I wish Dr. Panse speedy recovery from his recent illness and hope that his mature advice will be available to agricultural statisticians for many years to come.

(B. Sivaraman)

Dr. B. P. Pal
M Sc, Ph. D (Cantab) FLS, FRHS, FBS, FNI,
Director-General, I.C.A.R. and
Additional Secretary to the
Government of India
Ministry of Food & Agriculture
(Department of Agriculture)

Telephone :
Telegram : AGRISEC
भारतीय कृषि अनुसन्धान परिषद
कृषि भवन, डा० राजेन्द्र प्रसाद रोड, नई दिल्ली
INDIAN COUNCIL OF AGRICULTURAL RESEARCH
Krishi Bhawan, Dr. Rajendra Prasad Road,
NEW DELHI-1

Jaunary 1, 1967

I have known Dr. V. G. Panse over the last two decades. I have great regard and admiration for him as a scientist. His contributions in the field of agricultural statistics and allied subjects are well-known. His work has brought considerable credit to the Indian Council of Agricultural Research and has helped the Council in developing a strong centre for statistical research and training in the shape of the institute of Agricultural Research Statistics. The Indian Society of Agricultural Statistics of which he is a founder member has prospered under his able guidance. I am happy that the Society is felicitating him on his sixty-second birthday and has brought out on this occasion a special volume in his honour. He has my best wishes.

B.P.Pal
(B. P. Pal)

Telegrams : UNISERCOM

Chairman
UNION PUBLIC SERVICE COMMISSION
Post Box No. 186

New Delhi, the 3rd January, 1968

MESSAGE

I am very happy to learn that the Indian Society of Agricultural Statistics has decided to honour Dr. V. G. Panse by presenting him a volume of contributed papers from eminent scientists on the occasion of his sixty-second birthday. Dr. Panse served the Society for 16 years. I had the honour of knowing him since 1951 when he was appointed Statistical Adviser to the Indian Council of Agricultural Research. Agricultural statistics were improved considerably by the Indian Council of Agricultural Research, who did pioneering work first under the able guidance of Dr. P. V. Sukhatme and later of Dr. V. G. Panse in estimating the yields of agricultural crops, which until the Council took up this work were estimated by the Patwaries purely on the basis of visual inpection. Dr. Panse was also responsible for the starting of the Institute of Agricultural Resrarch Statistics, where training is given to students in the specialized subject of Agricultural Research Statistics. In addition, Dr. Panse had also occasion to advice the Food and Agriculture Organisation of the United Nations on various aspects of agricultural statistics. Dr. Panse retired from active service on attaining the age of superannuation and I hope he will have many years of useful retirement during which he will be able to give the benefit of his experience and knowledge to the present and the next generation of students of agricultural statistics so that the work which was initiated as well as encouraged by him during his tenure of office could be carried on further.

KRDamle

(K. R. Damle)

D. O. No.____________

M. S. RANDHAWA
D.Sc., F.N.I., I.C.S.,
Chief Commissioner,
Union Territory,
Chandigarh.

January 11, 1968

MESSAGE

I am very glad to learn that the Indian Society of Agricultural Statistics, New Delhi, have decided to honour Dr. V. G. Panse on his 62nd birthday by presenting him a volume of papers contributed by eminent scientists. I have known Dr. Panse ever since he joined the Indian Council of Agricultural Research as its Statistical Adviser. Dr. Panse is one of the most eminent statisticians of India and has a background of agriculture as well as agricultural research. I had the pleasure of knowing him and workimg with him when I was Vice President of the Indian Council of Agricultural Research from 1955 to 1960. I was greatly impressed by his intelligence and clarity of mind. His views on agricultural problems confronting the country were sound. By his studies he clearly stated the importance of chemical fertilizers in crop production. I only wish that his advice as well as of others who had similar views had been accepted in time by the Government of India. If it had been done the country would not have been in the present predictament. Dr. Panse is not only an eminent scientist but also a great administrator. Under his leadership the Institute of Agricultural Research Statistics made great progress.

I wish Dr. Panse many happy returns and convey him my hearty greetings.

(M. S. Randhawa)

FOREWORD

It gives me great pleasure to write this foreword to the Felicitation Volume to be presented to Dr. V. G. Panse by the Indian Society of Agricultural Statistics. The contributions of Dr. Panse in the field of Indian Agricultural Statistics and his labours in building up of the Indian Society of Agricultural Statistics are being very properly recognised by the presentation.

Almost the entire active work of Dr. Panse was carried out in Institutes of Agricultural Research. The broad field of his work was, therefore, the development of statistical methodology in relation to agricultural research. He specialised in a number of subjects in this broad field; particularly important was his work in some subjects of special interest for statisticians, such as sample surveys for estimation of crop yields, design of agricultural experiments and Statistical Genetics. In recent years Dr. Panse did a lot of work on subjects of practical operational interest, such as, application of statistical methods in demonstration and extension work. For economists, his cost studies have special interest. He took a leading part in designing the study in relation to comparative merits of the survey and the cost accounting methods in Farm Management Studies. I found his work in the field of Food and Nutritional requirements in India of particular value.

The high quality attained and maintained by Dr. Panse in his scientific work is attested by the large number of fellow scientists who have contributed articles to the Felicitation Volume. It is evidence of the influence exercised and the friendly admiration evoked by Dr. Panse that colleagues who had worked with him in a variety of contexts over the years should have readily agreed to join in this felicitation venture.

I have personally known Dr. Panse for many decades and have had the pleasure of working with him on a number of Committees. The strongest impression left on my mind is that of the great rectitude and the concern for scientific standards and values of Dr. Panse. These qualities are specially notable in one who has worked for almost all his life in official organisations and not in academic institutions. It would not be surprising if what

EDITORIAL NOTE

The Indian Society of Agricultural Statistics owes its present position largely to the vision, initiative, devotion and untiring efforts of Dr. V. G. Panse. It was, therefore, but appropriate that the Executive Council of the Society resolved to honour him for the excellent work done by him for the Society as Editor for a very long time and its Secretary till he retired from active work in December, 1966. It was decided to felicitate Dr. Panse on his sixty second birthday by presenting a Volume of papers in the fields of Statistics and Agricultural Sciences with which he was intimately connected.

The preparation of a Volume at a short notice was a stupendous task. There were only four months in which the volume had to be prepared. Dr. Panse has a large number of admirers and co-workers who would like to pay homage to him by contributing an article to the Volume. The choice of the subjects and the contributors presented a difficult problem to the Editorial Board. It was ultimately decided to leave the choice of the subject to the individuals, but the contributors were chosen to represent a wide spectrum of fields related to Dr. Panse's work. A list of persons from whom there was a likelihood of setting contributions for the volume in a short time was quickly prepared. For any errors of omission made in choosing such a list we aplogize to all concerned and not the least to Dr. Panse himself. Many more would have liked to contribute had there been no such limitation of time. It is indeed gratifying that almost all the persons who were invited to contribute did respond enthusiastically. Only a few declined the invitation due to previous commitments. To all the contributors who have made this Volume possible as well as to the other eminent persons who have sent messages to the Volume, we are very grateful. In particular we would like to express our deep gratitude to Prof. D.R. Gadgil who, in spite of his onerous duties as Deputy Chairman of the Planning Commission, spared time to contribute the preface to the Volume.

might have been considered as the unbending nature of Dr. Panse and his tenacious hold of principles and formulations which were correct according to his lights led him into difficulties in progress through official career. But the strength of character and the quality of his work were such that he was usually able to surmount all such difficulties. Therefore, entirely apart from the value of the particular contributions to statistical science of Dr. Panse, I consider the example set by his life and work as of particular importance. The demonstration that even within the official frame the scientific attitude and approach can be scrupulously guarded if the individuals value them with sufficient intensity and are ready to take some risk, is of great national importance.

I join with all the contributors to the Volume in the felicitations to be conveyed to Dr. Panse and hope fervently that he will be spared in good health for many years to give, with his work and example, guidance and inspiration to the younger generation.

D. R. GADGIL

CONTENTS

Our thanks are due to Shri K.C. Raut, Joint Secretary of the Society, who bore the brunt of the task of preparing the material for the Press and to Sarvashri A.K. Banerji, R. Remakanthan and P.C. Gupta of the I.A.R.S., who assisted him in this task. Our thanks are also due to Prof. G.S. Baird of the Rockefeller Foundation who spared us the stenographic help of Shri Neelakanthan and Shri V.P. Wadhwa of the I.A.R.S. who together with the former person undertook the painstaking task of accurate typing of the entire material within a very short time.

We are thankful to Messrs R.K. Printers for bringing out the Volume within a remarkably short time.

G. R. SETH

ON CONTRIBUTIONS OF VINAYAK GOVIND PANSE

by

P.V. SUKHATME

F.A.O., ROME

1. Introduction. The retirement of Dr. V.G. Panse from active work marks the end of an era in the development of agricultural statistics in India. He had just commenced his work as Adviser to the Planning Commission when illness forced him to retire in September 1966. Previously, he was Statistical Adviser to the Indian Council of Agricultural Research and Head of the Institute of Agricultural Research Statistics, which position he held until April 1966. He was a founder member of the Indian Society of Agricultural Statistics of which he had been Secretary since 1951. It is only natural that his friends and colleagues should wish to express their appreciation of his work by presenting him with a book incorporating articles in the fields in which he was most active. The task assigned to me on this occasion is to write a bibliographical note setting out a brief resume of his contributions to the development of agricultural statistics. I do so with the utmost pleasure and in so doing wish to offer my felicitations on his 62nd birthday.

2. Statistical Method in Agricultural Research. Born in Maharashtra on 11 January, 1906, Dr. Vinayak Govind Panse received his early education at Nasik and graduated in 1927 from the University of Bombay with mathematics, chemistry and physics as his subjects. He started his career as Agronomy and Chemical Assistant at the Institute of Plant Industry at Indore in 1927. It was here working with Dr. Hutchinson, the noted plant breeder, that Panse developed his interest in the use of quantitative techniques in agronomy. Plant breeding in those days was more an art than a science depending for the most part on the judgment and skill of the plant breeder. Mass selection, which the method in use and which consisted of choosing from the material under selection a number of plants which appeared to have superior value, bulking the seed from these, raising from the seed the next generation and continuing selection in this generation, was obviously inefficient in that the selection was subject to a large amount of environmental or non-genetic variability present in the field. Instead of bulking the seed obtained from different selected plants, breeders saw the need to sow it in separate progeny rows and to make the selections on the basis of progeny means. The selections thus made would be subject to only a fraction of the environmental variation to which the individual plants are subject. Nevertheless, there was no assurance in the method that the estimates of progeny means

would be unbiased. The small and variable amount of seed presented further difficulty in arranging replicated trials which could provide sufficiently accurate estimates of the progeny means. The difficulties in conducting randomised replicated tests were particularly large in the early stages of breeding when the material was even more heterozygous as the genetic values contributed to experimental error. Hutchinson and Panse set about adapting the randomised block and split plot designs to the plant breeding material at Indore and succeeded in developing what is known as replicated progeny row and compact family block designs. The results of their work were published in the Indian Journal of Agricultural Science (1935, 1937).

This, however, was only a beginning of more significant contributions to statistical methods in quantitative genetics. For, apart from an efficient method of field testing, a factor of fundamental importance in plant selection is the recognition of the fact that only a part of variability in plant population is genetic, the rest being environmental. The breeders' success depends upon the initial choice of material containing a high degree of genetic variability and exploiting it to the maximum extent possible. Panse (1940) showed how the genetic component of observed variability could be estimated by taking the regression of progeny means on parental value and explained the importance of selecting plants on their deviations from plant means rather than on their own values. He took the study of quantitative techniques a lot further by introducing appropriate genetic models which helped to bring out the effects of the number of segregating genes, the magnitude of their action, the modification due to dominance and the influence of environment on progress due to selection. This work which Panse carried out under the guidance of Sir Ronald Fisher in England was published in the Journal of Genetics (1940*a*) and Annals of Eugenics (1940*b*) and earned him Ph. D. from the University of London. In the normal circumstances, Dr. Panse would have continued this fruitful line of work but war interrupted his studies and he returned to India in June 1940.

On his return Panse was appointed to the newly created post of Statistician at the Institute of Plant Industry at Indore. It was then that I first came in contact with his work. I was statistician at the Indian Council of Agricultural Research in New Delhi and had just commenced a statistical analysis of the 10 years' data of a goat-breeding project at Etah in U.P. Drawing extensively on the methods developed by Panse, I found to my astonishment and dismay that the improvement in milk yield recorded in the goat breeding project at Etah was not so much due to a genetic improvement of the stock through selection as due to extraneous factors. Animal breeders naturally found it difficult to accept that a project run by an eminent breeder over 10 years could possibly have failed in recording a real improvement, particularly when year after year the reports showed appreciable increases in milk yield. However, the results of the statistical study left little doubt that the original stock did not have sufficient genetic variability ; and that in consequence no substantial progress could be achieved in improving the genetic potential of the herd for milk. But for the help in this work given by Dr. Panse, it would have been difficult to convince the animal breeders in those days of the need and value of statistical methods in planning animal breeding programmes and assessing their results. It was in fact the success of this statistical appraisal which led to the expansion of the statistical unit of the Indian Council of Agricultural Research (I.C.A.R.) which was later to form a full fledged Institute of Agricultural Research Statistics under the direction of Dr. Panse himself. The influence of Dr. Panse's early work can be seen in the increasing recognition, by animal husbandry workers all over the country, of the need for planning animal experiments based on statistical principles and for using statistical methods in the appraisal of their results. It led to important contributions from Panse himself and from his staff and students in the Division, notably from V.N. Amble, T.R. Puri and S.D. Bokil.

In 1946 Panse was promoted to the post of Deputy Director, Research, of the Institute of Plant Industry at Indore and later was made its Director. Panse's vision was so broad that no one could possibly have better filled this post which he held with such distinction until 1951 when he moved to Delhi as Statistical Adviser to I.C.A.R. It was during the years 1946-51 that I had an opportunity of frequently visiting the Institute of Plant Industry at Indore and to see at first hand the variety of experimental designs used on the farm, their simplicity and efficiency, and the use of statistical methods in the analysis of the simple experiments on cultivators' fields in districts around Indore which provided a wealth of information on response to fertilizers and other treatments with their sampling errors. It was discussion of the various problems of experimental designs and statistical methods which agricultural research workers faced in their work at the farm and in the cultivators' fields which led Panse and me to write jointly a book on Statistical Methods in Agricultural Research which continues to be the leading college textbook all over the country. The book was published by I.C.A.R. in 1954 and has already gone through three editions. It was translated into Spanish by Miss Ana Maria Flores and published by Fondo de Cultura, Mexico. It was mainly the work of Dr. Panse and bears the stamp of his practical knowledge acquired in the field.

3. Sampling Techniques for Estimating Yield. The Statistician at the Institute of Plant Industry was also the Statistician to the Indian Central Cotton Committee (ICCC) and in this latter capacity Dr. Panse was asked in 1941 to make proposals for improving the statistics of cotton production. Given adequate supervision, the ICCC saw no particular difficulty in obtaining reliable statistics of acreage under cotton through existing machinery in the cadastrally surveyed areas of the country. What the Committee wanted was a sampling method of objectively estimating the yield per acre in place of the subjective method in vogue in the country.

The ICCC had before it two schemes of research for estimating yield rates, one by Panse and the other by Prof. Mahalanobis. In his scheme Panse emphasised that any sampling method must fit into the existing administrative structure and take cognizance of the fact that the departmental staff and the farmers were already familiar with crop-cutting procedures. Mahalanobis, on the other hand, following earlier work by Hubback came forward with a scheme involving experiments with plots of small sizes—three or four concentric circular plots with radius of 2′, 4′, 5′, and 7½′. The two schemes were considered by the ICCC at a meeting of experts in 1942 which I had the privilege to attend. It was there that I first saw Panse putting forward justification for his scheme, not only based on sound statistical methodology but, what was more important, from an intimate knowledge of the difficulties one has to face when dealing with the cotton crop. The ICCC approved Panse's scheme. The approach was tried in one district to start with, and extended to two more districts in the subsequent year. The results were checked against the amount of cotton which came for ginning into the mills after making appropriate allowances for the import and export out of the selected districts. The latter data were gathered from officers specially posted for the purpose on the borders of the districts. The ICCC was satisfied with the results, approved the approach and came out with a recommendation in favour of asking the State Governments to adopt the method developed by Panse as a permanent solution for improving the forecasts of the production of cotton in the country.

It was not, however, the statistics of the cotton crop alone which benefited from the work of Panse. The method of crop-cutting surveys developed by him had a tremendous impact on the work which I.C.A.R. was asked to undertake at the time for improving statistics of acreage and yield of principal crops all over the country. So rapid was the progress of I.C.A.R. work that yield surveys based on the method became an annual routine in most States within the course of a few years. The surveys provided not only the State yields with a high degree of precision but also district yields with a reasonable margin of error. In 1951 Panse took over the direction of I.C.A.R. work himself and set out to extend the sampling technique to other crops, but the transfer of this work to National Sample Surveys (NSS) in 1952 blocked further progress. Panse thereafter devoted his energy and initiative to the development of sampling techniques in other agricultural fields, which we shall discuss later.

It must not be inferred that the progress made in estimating yield rate was easy to achieve. In particular, this work remained a matter for continuous criticism and opposition from Prof. Mahalanobis. Panse always believed that scientific work to be progressive must evoke criticism but criticism voiced through normal channels in scientific journals is one thing and criticism made through the channels of foreign expertise is another. Some of the foreign experts ostensibly invited for the purpose of lecturing in statistical theory, were deliberately involved in the controversy over matters of which they themselves had little firsthand knowledge. The Government itself in those days, it must be said, had a weakness for the advice of a foreign expert. The story of those days will perhaps never be known to most statisticians. Fortunately for India's statistics, Panse was persuaded to remain at his post but the eventual transfer of the work to NSS in 1952 deprived I.C.A.R. of the opportunity to gain from his ability and initiative in placing agricultural statistics of yield rates on a sound basis. The truth of this statement is brought home when one sees that India's statistics of crop yields remain much at the stage where they were left in 1952 with little or no further progress in the use of objective sampling methods by the State Governments.

Space does not permit me to deal at any length with even the basic issues of this controversy but no bibliographical note on Panse's contributions can have even a sense of completeness if I fail to mention here the way Panse felt about the whole problem. Perhaps the best way of doing this is to quote a few extracts from his own writings. This is how, for example, Panse summarised his approach to the method of estimating yield (IJAS, 1948, 1951).

"An examination of the above methods of estimating crops yields by random sampling, first tried by Hubback and later elaborated by Mahalanobis, leads to the conclusion that a new approach to the problem is needed if objective methods of crop estimation are to be introduced successfully in India. We have to take into account the existing administrative machinery and the fact that experimental harvesting of sample plots or crop cutting experiments is already a familiar routine both to government officials and farmers in different States. Insistence on a very small size of the plot cut with a special apparatus and requiring a delicate technique, in place of the large plot marked by chains or measuring tape and pegs, emphasises an altogether subsidiary aspect to the problem. Employment of special moving staff required for conducting the survey by these methods would involve heavy additional expenditure to which administrations are naturally reluctant to commit themselves permanently, when departmental crop cutting experiments are being carried out by the usual staff as part of their normal duties. Crop yield surveys in Bihar and Bengal conducted by Mahalanobis have demonstrated that an ad hoc staff of travelling investigators is unsuited for crop cutting surveys. Sampling methods for estimation of crop yields to be acceptable to State Governments should, therefore, avoid any radical departure from the established departmental procedure for crop cutting experiments, by making only essential changes to start with and the plan should be developed in such a manner that its permanent adoption will fit into the existing administrative structure and not necessitate any heavy additional expenditure on field staff. Once a start is made, however, gradual improvements in the design can always be introduced with the progress of work. These are the considerations that have guided us in developing the methods of random sampling-surveys for estimating crop yields, which we propose to describe in the present article.

To introduce a random sampling method for estimating yield, the only major change required in the traditional methods of harvesting sample plots is to ensure the selection of the plot for harvesting by a process of randomisation in place of the subjective selection by the experimenter. The problem is one of primarily convincing the administrator that a random selection of plots for cropcutting is both essential and practicable in the hands of the existing field agency. The number of plots to be harvested and their distribution over the tract as also the size of the sample plot are, however, matters to be settled by suitable experimentation. It follows that the yield survey should be so planned that while it fulfils its immediate objective, which is to give reliable estimates of yield for the tract surveyed, it should also furnish simultaneously technical information calculated to improve the efficiency of future surveys, so that yield estimates may be obtained with the requisite level of accuracy with a minimum expenditure of time and money and by maximum utilisation of available resources in the form of staff, ancillary information, etc."

What are the features of Mahalanobis' method ; how did Panse evaluate them and come to formulate his own approach ? The following extracts from the same papers in the I.J.A.S. (1951) are noteworthy for the way in which he dealt with the question.

"The fieldwork is entrusted to ad hoc parties of investigators who are required to move rapidly from place to place during the harvesting season to cut sample plots. The sampling is confined to those fields which are ready for harvest on the date of the investigator's visit to a place. A consequence of this method is that there is no guarantee that fields maturing at different times will be sampled for harvest in the proportion in which they occur in the population. The principle of randomisation requires that every field bearing the particular crop shall get due chance of being included in the sample. To achieve this objective, it is not only sufficient to select fields with the help of random numbers but it is equally important to ensure that there are no subsequent omissions of fields with certain characters correlated with yield such as fields in which the crop was already harvested before the experimenter could reach the spot or those where the crop was not ready for harvest at the time of his visit. Such deviation from randomisation will introduce bias and result in an under or over-estimation of the average yield—according to whether the rejected fields were better or lower yielders than the rest. As a general rule, late maturing crops are more vigorous and yield more under normal conditions than early maturing ones. Applying this test to the surveys, proper randomisation is lacking. Rejection of fields for reasons given above was allowed and appears to be common, as investigators were supplied with lists of randomly selected fields containing many more fields than were to be actually sampled as a precaution against cases of getting fields with no crop at all or fields which had already been harvested. This difficulty and the consequent possibility of biased results has been recognised by Mahalanobis from the beginning, but is inevitable in his plan of fieldwork. The difficulty of selecting plots for sampling was mentioned in the report of the first crop cutting experiments on jute in 1939 and as an example it is stated that in one village 182 plots had to be examined to secure the required 23 plots for sampling. This difficulty is again referred to in the Bihar survey. "After struggling with the problem for many years," Mahalanobis (1946*a*) concludes that "it is becoming clear that crop cutting work to be done properly must be carried out by a comparatively larger number of investigators who could watch the crop as it grows and collect sample cuts at the right time from the fields situated in the neighbourhood of their normal places of residence."

"The results of the Bengal crop survey also serve to emphasise the unsuitability of this approach to the problem of estimating crop yields. Out of the 7,680 sample cuts each aimed at for jute and aus paddy in the Bengal survey, only 1,190 cuts for jute and 2,028 cuts for paddy were actually secured. These represent about 15 per cent of the cuts planned for jute and 25 per cent for paddy. The reason given is that as the investigators had to move from one block to another, they often found that a considerable portion of the crops had already been harvested. This means that the crop actually sampled was mostly in the late maturing portion of the total crop. The extreme seriousness of the risk involved in estimating production of food crops on such data is obvious. The situation in the aman paddy survey was better in that about two-thirds of the sample cuts planned were actually harvested although even this proportion is far from satisfactory for giving reliable and unbiased estimates of yield. Mahalanobis' (1948) own comment on these results is a repetition of his remarks on the Bihar survey : "The crux of the whole matter is that it is essential that investigators should watch the crop as it grows and cut it as soon as ready for being harvested"; but he has now come to the conclusion that the solution he had proposed earlier, *viz.*, employing a larger number of investigators, was not practical on account of the expense involved. He adds : "Unfortunately, crop cutting work cannot be done by an ad hoc staff recruited merely for this purpose ; because careful selection of personnel is essential and the staff has to be given proper training as otherwise the work would be too unreliable to be of any value. One way would be to have a sufficient number of additional hands provided in the area survey scheme who would be available for crop cutting work at the proper season. This, however, was not feasible within the sanctioned grant. The crop cutting work had to be done, therefore, by the area survey staff as best they could manage it." One solution of this difficulty is to entrust crop cutting work to land records or agricultural agency as part of their normal duties, which precisely is the basis of our own approach to this problem."

Any discussion of the technique of crop estimation would be incomplete without a reference to the aspect concerning the size and shape of plot in crop cutting surveys. Experiments carried out by I.C.A.R. (1943, 1945, 1947) had shown that plots of small size, such as used by Mahalanobis, give an overestimation of yield. This bias resulting from plots of very small size at least under Indian conditions is recognised by all workers but there is no unanimity of opinion concerning the stage at which it becomes negligible. Of this Dr. Panse wrote in 1948 what is pertinent even today.

"There can be no difference of opinion that with training and supervision the limit at which bias becomes negligible may be reduced. In our view, however, it is not enough to demonstrate that plots of size as small as 50 sq ft. give unbiased estimates when handled by trained statisticians, nor is it adequate to say that very small plots such as circles of radius 2′ should give much less bias than that observed in our investigations. Such results or comments, in our view, have limited value for practical application in India where the employment of a special agency of statisticians for field work in crop sampling work is out of the question and experiments have necessarily to be carried out by the local experimental staff in course of their normal duties. In recommending a method for routine adoption it is therefore of utmost importance to ensure that the method is not susceptible to obvious source of bias in the hands of the agency handling it."

Plot size continues to remain a subject of active research in the country even today. It has become one of those prestigious questions for which funds will always be found irrespective of the need and significance of this research for immediate issue before the country. One only hopes that the practical side of the question so well stressed by Panse in the above quotation will not be forgotten in further research of the problem.

4. **Area Statistics.** When ICCC in 1942 asked Dr. Panse to make proposals for improving statistics of yield rates it saw no particular reason to question the accuracy of the statistics of acreage under cotton in the cadastrally surveyed areas of the country. There is a detailed and accurate frame available in these areas in the form of cadastral acreage maps. Simple objective methods of enumeration are employed. The data is collected in the course of normal administration by the local patwaris by plot to plot enumeration, which ensures their completeness, and there is provision for adequate supervision. All the same, at a later date, the Committee asked Dr. Panse to investigate whether the method worked efficiently in actual practice. Since it was considered possible that the work might suffer from lack of attention, both from the primary agency and the supervisory staff. The main purpose of the enquiry was to ascertain whether the burden of work involved was excessive to be carried out in the course of normal work by the Patwaris and if there was any possibility of reinforcing supervision on a rationalised basis using sampling methods for the purpose. Short of appoint-

ing special field agencies to collect agricultural statistics, which luxury even the advanced countries could not afford, the Committee thought that this was the only way in which the required improvement, if any, in the collection of area statistics could be brought about.

Panse carried out a number of sample surveys to check the accuracy of area statistics and brought the results together which showed that the method of plot to plot enumeration in the cadastrally surveyed areas with Patwari agency worked satisfactorily in practice. Nevertheless, criticism continued to be made about the Patwari system and the reliability of the acreage statistics obtained by them. Some critics even went to the extent of condemning the system as being incapable of providing reliable statistics. The principal critic was Prof. Mahalanobis. Using the field agency of the National Sample Survey, he conducted a check on the work of the Patwaris in four northern States of India in 1949 and 1950 and showed that Patwaris overestimate acreage under wheat to the extent of 10-26 per cent in the different States with an average overestimate of 15 per cent. On the other hand, cash crops like sugarcane, linseed and other oil seeds were shown to have been underestimated, usually more than 50 per cent. The results of this enquiry caused concern in the Government of India. Careful examination, however, of the results showed that the discrepancies between the Patwari records and those of the NSS investigators were due to *the differences in concepts and definitions used by the two agencies*. Sugarcane, for example, although planted in the months of January to March, is entered in the Patwari register only at the time of the crop inspection during September/October. This is in accordance with the Land Records Manual of Instruction for recording crop acreage. The NSS investigator on the other hand, noticing the crop in the field during his inspection in April and finding no entry by the Patwari recorded it as a mistake. Again, linseed and other oil seeds grown mixed with wheat and other cereals are not recorded, the entire area being shown against the cereals. This is in accordance with the Land Records Manual. At the time of final computation of crop acreage, however, an appropriate allowance is made in the area under cereals for this omission. These facts were found to account for the observed discrepancies in acreage under wheat as well as oil seeds. Clearly, the sample check conducted by Prof. Mahalanobis was not valid in critically assessing the efficiency of the administrative enumeration procedures.

A number of independent sample checks were subsequently conducted by Panse himself which confirmed that leaving aside questions of concepts and definitions, the acreage statistics derived from the Patwari census were on the whole satisfactory. All over the world in many advanced countries statistics of crop acreage are in fact collected by annual census through the available administrative machinery, such as village or municipal committees. Human nature is also the same everywhere. It is ridiculous to suggest that an agency belonging to one department is more trustworthy than that belonging to another. As Panse put it, the integrity of our statistics like all our activities will develop with our national character. Given adequate supervision there is therefore nothing in the census method of plot enumeration in cadastrally surveyed areas to doubt the accuracy of acreage statistics. Panse was therefore rightly concerned that the country should spend its energies in condemning the Patwari system instead of attempting to strengthen it and to extend it to unsurveyed areas where it did not exist. In any case, when the data is needed for very small administrative units as a basis for regional planning or in order to provide benchmark information for current agricultural statistics, sampling methods may be uneconomic and complete enumeration inevitable except for items which do not lend themselves to complete enumeration. This is not to say that Panse did not recognise the role of sampling method in improving acreage statistics. Indeed, in his view, sampling had an indispensable roll in controlling and improving the quality of fieldwork of the primary enumerators and in speeding up the availability of the results of the census. In particular, sampling had a great role in rationalising the supervision of Patwari work and in improving the reliability of early forecasts of crop acreage. A coordinated scheme for the purpose was drawn up by I.C.A.R. in agreement with the States but after transfer of the work to NSS it was not possible to implement it.

Panse also welcomed the use of sampling method for areas which were not cadastrally surveyed, which did not have a Patwari agency and which, in consequence, could not possibly use a census method of plot enumeration. Indeed, he himself gave a lead by organizing an area survey based on sample method in the unsurveyed State of Orissa. However, to condemn what had been developed over decades in the States having Patwari system, without trying to improve it with the help of sampling methods, such as randomised supervision, was to waste the limited sources available for the improvement of statistics. It was his hope that the NSS Organisation would exercise a healthy influence on the Patwari system by supplementing the efforts of the State Gevernments by providing the needed supervision on a sampling basis. With its federal system of government and the main responsibility for agriculture lying with the States, he maintained that what India needed most was agreement on methods of collecting statistics using available field agency reinforced by supervision by the centre on a sampling basis. If NSS could only achieve this, a great step forward would have been taken in improving the much needed statistics of production for facilitating the food administration of the country.

5. **The Census of Agriculture.** Panse made a significant contribution to the programme and progress of the decennial census of agriculture sponsored by FAO. Realizing that lack of basic data constituted a serious handicap to most of the developing countries in the preparation of their development plans, Panse saw in the census the first step in planning. At the same

time he kept in mind the possible impact of census on the evolution of a sound and permanent system of current agricultural statistics. For two years, 1960-61, he worked as Regional Adviser for Agricultural Census in Asia and the Far East, visited a number of countries in the region to observe the agricultural census in operation and discussed the problem of census-taking with field workers, directors and other technicians engaged in planning, organising and conducting the agricultural census. He incorporated this experience and his thinking on the subject in a manual he wrote for FAO on the problems of agricultural census taking, with special reference to developing countries. In this manual Panse extended the concept of the agricultural census itself as an integrated system of surveys. The manual is available in all the official languages of FAO.

It would be rash to conclude that Panse favoured complete enumeration to sampling method for collecting information on items other than land use in the cadastrally surveyed areas. I cannot do better than quote his views from the same manual on agricultural census taking. Panse wrote :

"A view is sometimes expressed that an agricultural census by complete enumeration is a desirable undertaking even if the results are likely to be of imperfect quality and have a limited scope. This view, however, cannot be accepted lightly, for a supposedly complete enumeration census which is in reality substantially incomplete, would produce results of poor quality and is liable to present a distorted picture of the agriculture structure if carried out under conditions of difficult communications, an ignorant and suspicious peasantry providing biased information and an unmanageable larger number of ill-qualified and ill-trained field staff that must necessarily be employed. Non-sampling errors, such as biased responses, are likely to be appreciable even where the census is substantially complete and it may turn out that the results are consequently less accurate than those of a well-conducted sample census subject to reasonably small sampling errors. It is against this background that a decision whether to carry out a complete enumeration or to project a sample survey will need to be taken. It will also depend upon the level at which the results are required, that is, whether the results will be tabulated for the entire country, for individual provinces, for individual districts or even for their administrative subdivisions. Of course, this decision need not be either a complete enumeration or a sample survey as the only alternative but could well be a combination of both."

6. Extension of Sampling Techniques to Other Fields. As mentioned earlier, Panse extended the application of sampling techniques to a number of other fields. Cost of production studies, estimation of catch of fish, estimation of livestock and their products, and evaluation of the intensive agricultural district programmes are a few examples of these fields. Space does not permit us to describe these applications beyond mentioning principal features of one or two of them.

The analysis and control of respondent bias were the principal features of his studies on cost of production. Results of the investigations carried out in different regions of India for three years under the auspices of the Research Progress Committee of the Planning Commission to compare two methods for collecting data for estimating cost of production on crops have confirmed this need for controlling non-sampling errors. One method consisted in locating an investigator in a selected village on a whole-time basis to record on sample farms field operations and other items by daily physical observation. In the other method the investigator paid periodical visits, three or four times a year, to selected villages and interviewed farmers of selected holdings to obtain data for cultivation and other operations since his previous visit. The investigators who did the field work were trained men with an agricultural background. The investigators employed for the interview method had an additional qualification in agricultural economics. The interview method gave appreciably higher estimates of total inputs and components like human and bullock labour, and underestimated the items of total output. The conclusion was that when ad hoc staff visit the selected farms periodically and obtain information by interview the input factors are inflated and output factors are underestimated. However, when the field agency is stationed on a group of farms and watches the operations over the season, this bias largely disappears. The results of these studies are brought together in a book entitled *The Techniques of Cost Production Studies* which illustrates in a remarkably lucid style the need and importance of controlling non-sampling errors in surveys.

Fish catches are made all the year round by almost innumerable small units operating along the sea coast of India. The illiteracy and ignorance of the fishermen of the quantitative aspect of their work and ingrained suspicion of any enquiry make the data on catches, secured through verbal enquiry, too unreliable to be of much use. Complete enumeration under these conditions is ruled out as impracticable. The total amount of the fish catch has, therefore, to be estimated by sampling in both time and space and employing objective method of enumeration, *viz.*, as physical measurement of the sampled catch. The solution to the various problems in sampling theory and practice raised by the enquiry makes one of the most fascinating readings and I can do no better than refer the reader to the original report (I.C.A.R., 1950) and the subsequent publication of the results in *Biometrics* (1958).

So interesting and novel was the application of sampling technique to fishery catch that FAO invited Panse to visit a number of countries for advising them on the improvement of fishery statistics of catch. In particular, Panse assisted the Governments of the United Arab Republic and Uganda in evolving an appropriate technique for the purpose. The results of the work are described in the report published by FAO (1964).

7. Other Contributions to Economic And Planning Statistics. Panse's interests have been much wider. His papers published in the Indian Journal of Agricultural Economics on the index numbers of agricultural production and on trends of crop yields provide two illuminating examples of his interest in developing procedures for evaluating the progress of development plans in the country. The latter paper in particular illustrates in a striking manner the contribution which statistical method can make to the development of evaluation procedures. It is common knowledge that comparison of annual yields suffers from seasonal variation. The difficulty is sought to be got over by comparing the average yields over several years. Even so, it is not possible to rid the average yields completely of seasonal disturbances, such as rust epidemic on wheat in consecutive years. There are a multitude of weather factors which influence crop yield and have to be allowed for in judging the progress of yield. Clearly, the total variation observed among annual yields for different districts in the country needs to be partitioned into variation (*a*) between the two plan periods, (*b*) between individual years within each period, (*c*) between districts, (*d*) representing interaction between periods and districts and (*e*) representing interaction between individual years within periods with districts. It is only after such partitioning that a comparison of components (*a*) and (*e*) will show whether the differences in the average yield level between the two plan periods is statistically significant or is only such as can arise from random fluctuation of seasonal yields and can therefore be reasonably ascribed to the latter. The analysis was without any effect on wheat but the adjusted rice yields showed an overall increase of 8 per cent during the first plan period as compared to the pre-plan period. Clearly, weather factors had adversely affected yield during the first plan period compared to the pre-plan period.

Panse's contributions to the development of yardsticks to assess the progress of agriculture have been particularly important. He brought together and analysed the massive data of field experiments on crop response to fertilisers and other inputs in a series of publications entitled *Index of Experiments*. These voluminous publications, one each for the different States of the country, were prepared under the guidance of an Expert Committee with Panse as Member-Secretary. They contain a wealth of information of great value for planning. It is this work which provided the basic data for developing yardsticks for judging the possibilities of increasing agricultural production in the country and enabled Panse to probe deeply into the technical aspects of India's plan for the development of agriculture. The assessment of the Intensive Agricultural District Program (IADP) in the country which he carried out for the Sen Committee of the Planning Commission provides an excellent example of this work by Panse.

His most recent article in June 1966 published in the Economic Times under the heading *The New Strategy in Agriculture* provides another example. In this article Panse scrutinised the claims of the new strategy in a manner which focussed the attention of the public and the Planning Commission alike. When the new strategy for the development of agriculture was announced in 1964, one was given to understand that one could expect an additional yield of one ton/acre on the average with a fertiliser dose of 100 lbs. of Nitrogen and 30 lbs. of Phosphate. Panse was all in favour of the new strategy but he did not find the available evidence justified the claim of the response of 1 ton/acre. With forthright candour characteristic of him, Panse showed that the available evidence did not justify an assumption of more than half the figure put forward by the Planning Commission. He feared that the tall claim of 1 ton/acre might lead to over optimism in the country and insistence on large dose may lead to disappointment considering that fertilizers were in short supply and a scarce resource and needed to be used at doses which gave the maximum additional response per unit of application. Without the support of essential scientific data the strategy was like a 'big lead forward' and he was afraid that unless appropriate steps were simultaneously taken it may meet the same fate. This article which Panse wrote in the Economic Times sometimes after he had retired from I.C.A.R., unfortunately lost Panse a good deal of his popularity in Government circles for the forthright way in which he expressed his criticism on the strategy. But that has always been the way in which Panse worked and wrote. He resented doing anything to please anybody, however highly placed. As the Vice-President of I.C.A.R. said at the farewell address, "Panse believed in the use of the hammer and not persuasion." Panse's reply was that his experience of over 35 years had taught him that persuasion led nowhere in India and it was his experience which taught him gradually harden in his approach and eventually in the use of the hammer. Panse never sought for himself any advantage out of any controversy. It has been my greatest privilege in life to enjoy his friendship. Not a month has passed when he and I have not exchanged views on some technical topic or another. May God give him the strength to recover fully from his illness.

STATISTICAL ASSESSMENT OF THE INTENSIVE AGRICULTURAL DISTRICT PROGRAMME

by

S.R. SEN

I. Introduction. I have had the privilege of working closely with Dr. V.G. Panse in the Ministry of Food and Agriculture in the Planning Commission, in a number of Statistical Committees set up by the Government of India and in the Indian Society of Agricultural Statistics, ever since he came to New Delhi in 1951 as Statistical Adviser to the Indian Council of Agricultural Research.

Of the various fields in which we worked together, special mention may be made of improvement of crop estimates, collection of farm management data, estimation of production functions, application of statistical methods to agricultural planning and statistical assessment of agricultural development programmes. In each of these fields, Dr. Panse made pioneering contributions of both fundamental and applied nature, about some of which an account has been given in other articles included in this volume.

I propose to confine myself to the contribution that Dr. Panse has made as the Member-Secretary of the Expert Committee on Assessment and Evaluation of the Intensive Agricultural District Programme, of which I happened to be the Chairman. It is to the work of this Committee that Dr. Panse devoted the last five years of his official career. He found special interest in this work because it gave him a unique opportunity to apply to a single programme in a coordinated manner all the expertise that he had acquired and the various techniques that he had developed in the course of a long professional career in the diverse fields mentioned in the previous paragraph.

The masterly way in which he tackled the difficult organisational problems of a statistical survey designed to help assessment and evaluation of an agricultural programme that covered at first seven and later fifteen districts scattered all over India was itself a great achievement. But this was not merely a survey on a very extensive scale but also of a very complicated and intensive character and gave rise to many difficult technical problems, each of which was successfully solved by Dr. Panse with his characteristic thoroughness and efficiency. However difficult the problem was, he never compromised his principles and was never satisfied with short cuts, sometimes advocated by his less intrepid colleagues, if these were considered by him to be technically not good enough. He struggled with the problem until he found a solution which was technically unimpeachable.

2. The I.A.D.P. Survey. The first seven districts with which the survey started in 1961 were Thanjavur (Madras), West Godavari (A.P.), Shahabad (Bihar), Raipur (M.P.), Aligarh (U.P.), Ludhiana (Punjab) and Pali (Rajasthan). They covered a total of 14,000 villages, 6·3 million hectares and 12·3 million people and produced as diverse crops as rice, wheat, jowar, bajra, maize, barley, gram, pulses, sugarcane, cotton, groundnut, potato, tobacco and chillies, each having its special problem of estimation. The agro-economic conditions also differed considerably from district to district and presented special problems of assessment. The survey covered not only physical data regarding inputs and outputs but also economic aspects like land tenure, credit, marketing, costs, benefits, agronomic practices, demonstrations, farmers' participation and other information needed both for the formulation of district and farm plans and their evaluation.

The first step was to organise a "bench-mark" survey for each of these districts which covered the various aspects mentioned above and gave an objective picture of the situation obtaining in the base period with which future achievements could be compared. In addition, for each district a "control area" was selected in a neighbouring similarly endowed district for which also similar data were collected for purposes of comparative study.

The "bench-mark" survey had to be completed in a relatively short time and in many places with inexperienced field staff. It also raised many technical problems, of both conceptual and operational nature, which had not been tackled earlier. Dr. Panse manfully overcame all these problems and completed the "bench-mark" survey according to schedule. The great emphasis that he put on and the careful arrangements that he made for the training of the field staff, preparation of operational manuals and effective supervision of the field and analytical work deserve special mention.

The "bench-mark" survey was followed by assessment and evaluation surveys every season in the "programme districts" as well as "control areas". They were also supplemented by special surveys, and studies were not only compiled and analysed for the published report of the Expert Committee but an attempt was also made to feed them back to the administrative authorities of the programme at different levels so that they could take advantage of the findings for improving the current operations. It must be admitted that the latter attempt was faced with considerable difficulty in practice and did not come up to the expectations. There are, however, signs of gradual improvement in the situation.

The concept as well as selection of "control areas" raised a number of controversial issues and difficult technical problems. Dr. Panse strongly pressed for the concept being accepted as a tool for objective assessment of the impact of the programme under study and also for making appropriate allowances for effects of extraneous factors like weather. All his expectations about "control areas" were not fulfilled. The main difficulties arose from the fact that "control areas" (which by their very nature had to be fairly extensive unlike the usual "control plots") could not be found which were exactly similar to the "programme districts" in the base period, and which could be kept entirely uninfluenced by the demonstration effects of the "programme districts" and where major structural changes could be kept in abeyance during the period of the study. However, the data that Dr. Panse collected and the analysis that he attempted about relative changes in "programme districts" *vis-a-vis* "control areas" proved quite useful in a number of cases. The conceptual and operational problems that he faced in this context and the way he tried to tackle them should be of considerable interest to all research workers in this field.

3. **The Design.** The survey comprised of two main parts: (*i*) agronomic and agro-economic enquiry in sample villages and for example cultivators' holdings within these villages and (*ii*) crop cutting experiments in important food and cash crops grown in the districts in randomly selected fields.

The design of the survey comprised a stratified multi-stage random sample. For the agro-economic enquiry, a zone consisting of 2 to 4 blocks constituted a stratum, and a village and a cultivators' holding in the village were the first and the second stage sampling units respectively. For the crop-cutting experiment, a block or a VLW circle was the stratum and a village, a field and a plot were respectively the successive stages of sampling units.

For the agro-economic enquiry about 800 cultivators' holdings spread in 100 villages were canvassed in each of the "programme districts" and about 300 holdings from 40 villages in each of the "control areas". The data was collected by an interview with the farmers, but a number of spot checks were also carried out.

For the crop-cutting survey about 300 experiments were conducted every crop season for each important crop in each "programme district" and 50 to 75 experiments on each of such crops in each control area. Yield data for these experiments were obtained by actual harvest and weighment of produce.

4. **Operational Problems.** The crop-cutting experiments were carried out following the well-known I.C.A.R. techniques, in the development of which Dr. Panse himself had made pioneering contributions earlier in his career. The results once again proved the efficiency and high degree of reliability of these techniques. An innovation was made to collect along with the data for output also the data for main inputs like fertilizers, pesticides, improved seeds, irrigation water, etc., for the same sample plots and their correlation and analysis and comparison with similar data from experimental and demonstration plots which proved highly useful. Some of

the techniques for determining production functions which Dr. Panse had experimented with earlier in the I.C.A.R. proved to be of considerable value in this context. These studies also threw interesting light on the divergences between production functions obtained in experimental stations and in the fields of progressive and ordinary farmers under Indian conditions.

The agro-economic surveys were somewhat novel and raised many difficult conceptual as well as operational problems. While Dr. Panse's earlier association with the Farm Management surveys jointly sponsored by the Planning Commission and the Ministry of Food and Agriculture gave him valuable insight into some of the problems involved, he had to think out several useful innovations in this special context. The techniques now developed as a result of these surveys have gone a long way to give us a much more reliable method of collecting data of this type than what obtained in India hitherto.

All these basic data and the various input-output co-efficients which were collected at Dr. Panse's initiative and stored at the Institute of Agricultural Research Statistics, proved very useful for both overall planning of the district programmes and formulation of the farm plans. In this context some of the techniques for agricultural planning which were evolved earlier by the Indian Society of Agricultural Statistics at Dr. Panse's initiative, proved to be of considerable value.

Thanks again to Dr. Panse's strong emphasis on statistical standards, the evaluation of the programme under reference has been much more objective than any other similar evaluation that I know of. Like a true scientist, Dr. Panse insisted on the exclusion of all comments which were not based on scientifically collected data and rejection of all data about the statistical validity of which there were doubts. Here again some of the techniques of statistical assessment, *e.g.*, for the spot check of the irrigation programme, which he had developed in the I.C.AR. earlier proved to be of considerable use.

The operational research type of study initiated under the programme was another innovation so far as Indian agriculture is concerned. Dr. Panse took a keen interest in this study and organised a special seminar and training course for this purpose at the Institute of Agricultural Research Statistics. These studies have no doubt been facing serious teething troubles and the progress has been much less than earlier hoped for. But I have no doubt that they are steps in the right direction and will yield rich dividends in the not too distant future.

5. The Results. Some of the main results of these surveys and studies have now been published in the three reports which have so far been issued by the Expert Committee. This is not the place to comment on these results. Whatever opinion one may have about them, one thing is beyond controversy. The statistics included in these reports are the most scientific of their kind collected in India so far.

But what has been published is only a very small fraction of what has been collected. Bulk of the latter have been appropriately stored in the Institute of Agricultural Research Statistics for the use of future research workers.

Both these data and the Institute will remain for a long time standing tributes to the great contribution that Dr. Panse has made to the science of statistics in particular and to his country in general in the course of a most fruitful professional career.

STATISTICAL QUALITY CONTROL APPLIED TO THE FOOD INDUSTRY

by

ANA MARIA FLORES*

Statistical Quality Control applied to Industry is a powerful tool in the industrial and economical development of a country.

This technique has been advancing very rapidly and is an important branch of Mathematical Statistics. The theory of Probability and the Sampling Theory are applied to save time and money.

When an industry is trying to measure the quality of its manufactured products, it is very likely that if it does not use Statistical Quality Control, it will confront a disagreeable situation; the product will not have the desired quality, a great deal of money will have been spent, there will have been an unnecessary waste of material and much time will have elapsed in its manufacture before it can be sent into the national or international markets. The product will have little acceptance because it lacks the required standards, because of its high price and because it is offered out of time!

The Statistical Quality Control is a technique that every industrial engineer should know so that it can be applied to the different steps of an industrial process. Its application is not difficult and if supervised by the engineer, any of the workers at the factory can apply it without neglecting his usual occupation and with the feeling that he is being more useful to the Company he works for, and even creating new ideas which can be very useful to the Industry.

There is a lot of literature on the Theory of Statistical Quality Control and it is not the purpose of this paper to call attention on this knowledge.

In this paper I propose to give a more advanced step ; by giving some ideas and formulae tending to help the food canning industries.

Canned foods are being sold very well the world over, but unfortunately in case the statistical quality control has not been enforced during the process of their manufacture, the foods sold may be substandard, thereby endangering the health and life of the consumers.

We should think of canned foods as a multi-assembled product.

---

*Ministry of Industry and Commerce, Mexico City—Mexico.

Let A be the canned product ready for sale, let $B_1, B_2, \ldots B_n$ be all the different steps taken in the manufacture of the product :

Let $X, Y_1, Y_2, \ldots Y_n$ be random variables related respectively to $A, B_1, B_2, \; B_n$ with normal distributions and with standard deviations

$$\sigma, \sigma_1, \sigma_2, \ldots \sigma_n.$$

We define

$$A = [\,|\,X\,| \;\; < 1{\cdot}96\;\sigma] \qquad \ldots(1)$$

$$B = [\,|\;Y_j\;| < 1{\cdot}96\;\sigma_j]\;;\;\; j = 1, 2, \ldots n, \qquad \ldots(2)$$

taking for t as the value 1·96 at a confidence level of 95 per cent.

Then

$$A = A' \cap \bigcap_{j=1}^{n} B_j \qquad \ldots(3)$$

because the event A consists of the intersections of the events A' and all the B_j and where A' is the event in which the normal random variable X', measures the quality of the assemblage of the B_j parts distance of which from the mean $\bar{X}$ is less than $1{\cdot}96\;\sigma'$:

$$A' = [\;|\;X\;|\; < \;1{\cdot}96\;\sigma']$$

From (3) we obtain

$$A^c = A'^c + \bigcup_{j=1}^{n} B_j^c$$

Where A^c is the complement of A, and then

$$P[\;|\;X\;| \geqslant 1{\cdot}96\sigma] \leqslant P[\;|\;X'\;| \geqslant 1{\cdot}96\sigma] + \sum_{j=1}^{n} P[\;|\;Y_j\;| \geqslant 1{\cdot}96\sigma] \;\ldots(4)$$

and now we can use a Theorem by Valle Flores*

Theorem : "If Y is a random variable and f, is an even function

$$f(-x) = f(x)$$

(Borel), not negative and increasing on $[0, \infty]$, then for every real number $\gamma \geqslant 0$ we have :

$$P[\;|\;Y\;| \geqslant \gamma] \leqslant \frac{E[f\;|\;Y\;|\;]}{f(\gamma)} \qquad \ldots(5)$$

where P is the probability of the even t and E is the expectation of the random function (Y)."

Then by (4) and (5)

$$P[\;|\;X\;| \geqslant 1{\cdot}96\;\sigma] \leqslant \frac{E[g(x')]}{g(1{\cdot}96\sigma')} + \sum_{j=1}^{n} \frac{E[f_j(y_j)]}{f_j(1{\cdot}96\sigma_j)} \qquad \ldots(6)$$

---

*See Boletin de Tecnicas Aplicaciones del Muestreo". No. 7, pp. 9-12.

where the g and f_j are the real functions of the real variables which satisfies the above theorem.

Now, we have to choose adequately the random functions g and f_j so as to set up in advance the control on the X' and Y_j (the quality control in the manufacture of the parts B_j and the assemblage of them), so as to guarantee enough quality of the industrial product A.

For example,

$PA^c \leqslant 0{\cdot}05$, or equivalently, $PA \geqslant 0{\cdot}95$

As an application to the above, let us suppose that we are interested in the fish canning industry.

Let us consider A as the lot of fish cans manufactured and ready for sale, and B_i the different steps taken for the canning of the product.

B_1 = Selection of the best fish from among all the available fresh fish in the factory.

B_2 = Cutting of the fish into pieces, of, say, one pound each,

B_3 = Taking out the bones from these pieces.

B_4 = Preparation of the tomato sauce.

B_5 = The pouring of the sauce into the fish.

B_6 = Treatment of the cans, such as sterilizing in boiling water and chemicals.

B_7 = Packing the cans with the prepared fish.

B_8 = Sealing the filled cans under pressure.

B_9 = Cooking the filled cans at high temperatures to avoid bacteria.

In this case we can see that not all the events are independent, and so, we need to use some formulae for the dependent events in order to calculate the total probability of obtaining a good reasonable level of quality in the final product A.

Suppose that we have the following probabilities in every step :

$$P(B_1) = {\cdot}995$$
$$P(B_2) = {\cdot}990$$
$$P(B_3) = {\cdot}985$$
$$P(B_4) = {\cdot}984$$
$$P(B_5) = {\cdot}959$$
$$P(B_6) = {\cdot}979$$
$$P(B_7) = {\cdot}999$$
$$P(B_8) = {\cdot}968$$
$$P(B_9) = {\cdot}999$$

ON TRENDS IN THE BOVINE POPULATION OF INDIA

by

M.S. AVADHANI[1], R. GOPALAN[2] AND V.N. AMBLE[2]

1. Introduction

In recent years Dr. V.G. Panse has taken increasing interest in the subject of planning for agricultural development. It was in pursuance of this interest that he suggested that the Research Unit of the Indian Society of Agricultural Statistics should take up as the first project, investigations into the building up of a suitable statistical model for livestock development. The study reported in the present paper is a part of these investigations. It seems appropriate to offer this as a token of the high regard with which the authors hold Dr. Panse.

Projection of livestock numbers on the basis of the available livestock data is fundamental to any investigation on setting up a meaningful statistical model for studying development of livestock. In general this problem can be studied by constructing life-tables for the population under consideration, or by determining trends in various categories of the population. Since birth rates and sex-wise age-specific death rates are essential for the construction of life-tables and these statistics are not available, the latter procedure is the only feasible alternative in the present case. The nature and quality of the livestock census data available until 1961 and the methods employed to utilize it for the study of trends in various categories of cattle and buffaloes are described in this paper. Projections of the bovine population worked out on the basis of these trends are also presented.

2. The Material

The nature and quality of the livestock census data available until 1961 and the procedure adopted to assemble them for the present study are described in this section.

2.1. **Nature of the Available Data.** The first livestock census was held in 1919-20 and until Independence it was carried out almost quinquennially throughout British India excepting in the Central Provinces and Berar where they were held annually. Although the native Indian States were invited by the then Government of India to hold the livestock censuses simultaneously, only a few States participated in the census right from the

---

[1] Indian Society of Agricultural Statistics, c/o I.A.R.S., New Delhi.

[2] Institute of Agricultural Research Statistics, New Delhi.

Then, if the first two steps are subject to a conditional probability, we shall have

$$P(B_1B_2) = P(B_2 \mid B_1)\ P(B_1)$$
$$= \cdot 990 \times \cdot 995 = \cdot 985$$

Then

$$P(A) = \cdot 985 \times \cdot 985 \times \cdot 984 \times \cdot 959 \times \cdot 979 \times \cdot 999 \times \cdot 968 \times \cdot 999$$
$$P(A) = 0 \cdot 87$$

which is the probability of the quality product A, the canned fish. Or also, the probability of A^c which is the possible error in the quality, will be $P(A^c) = 0 \cdot 13$, which means that it is likely that 13 per cent of the cans are not in perfect quality.

The functions $g(X)$ and $f_j(Y_j)$ are induced by the random variables X and Y_j and can be known when one takes samples of X and Y_j respectively.

As we can see from the example given, it is necessary to know and to measure the probability of the quality in every one of the steps taken, and to try to improve this quality in order to obtain the best quality possible in the canned product.

That is the reason why Statistical Quality Control must always be used in food industries.

beginning. Whereas some States such as Hyderabad, Mysore, Travancore-Cochin, Banganapalle, Sandur, Pudukkottai, etc., consistently participated in the census from the very inception, many other States, notably those in Central India, and areas now in Maharashtra, Gujarat and Punjab States did not participate in the census taking for a fairly long period. Bengal, Bihar and Orissa did not participate in the fourth census (1935) and U.P. in the fifth census (1940). Whereas the sixth census (1945) was generally conducted in all the Provinces and Administrations in British India, U.P. conducted this census in 1944 and Bengal in 1946. All the native States which participated in the livestock censuses held the censuses almost simultaneously with the British Provinces.

The period of enumeration was four months from December to April for the first two censuses, and it was one month, *viz.*, January, for the subsequent censuses held before Independence.

The seventh quinquennial livestock census due in January 1950 was postponed with a view to combine this census with the F.A.O. Agricultural Holdings census. Subsequently, however, when the latter was postponed (and was held afterwards on a sample basis), the livestock census was held in May, 1951, or later in all the States of the Indian Union excepting in Orissa and Manipur which could not participate in this census at all. Ajmer and PEPSU, West Bengal, Rajashan and Travancore-Cochin conducted this census in June, 1951, September, 1951, February, 1952, and August, 1952 respectively. The eighth census was held generally in all the reorganised States and Union Territories of the country in March-April, 1956, the reference date being 15th April, 1956. West Bengal, Orissa and Manipur however held this census in 1957, the reference dates being 15th April 1956, 15th April 1957, and 15th December 1957, respectively. The ninth livestock census was conducted in all the States and Union Territories of the country in 1961 with 15th April, 1961 as the reference date.

In all the six censuses held before Independence throughout British India and some of the erstwhile native Indian States information in respect of cattle was collected under the heads : (*i*) Bulls, (*ii*) Bullocks, (*iii*) Cows, and (*iv*) Young stock, and, buffaloes under the heads : (*i*) Male buffaloes, (*ii*) Female buffaloes, and (*iii*) Young stock. Cattle or buffaloes 'not old enough for work or breeding' were treated as young stock, and it was reported that livestock in cities and cantonments were 'included *wherever it was possible* to secure their enumeration'. It is obvious that such a classification of the bovine stock is quite ambiguous and the recorded census figures for these censuses would be consequently subject to non-sampling errors of unknown magnitude. It may also be pointed out that the effect of vague terminology used for the classification of stock is likely to be more pronounced on young stock owing to small numbers in this category than on the other categories.

Some major improvements were introduced in respect of the terminology and classification of stock used for the three censuses held in the country after Independence. In the 1951 census and the subsequent ones all bovine stock of age above three years were considered adults and those with age three years or below as young stock. For purposes of 1951 census adult males of cattle and buffaloes were classified into (*i*) Breeding bulls, (*ii*) Working bullocks, and (*iii*) Others, adult females into (*i*) In milk, (*ii*) Dry, (*iii*) Not calved even once, (*iv*) Used for work and (*v*) Others, and young stock were divided sex-wise into (*i*) Under one year and (*ii*) One to three years. Additional significant changes introduced in the 1956 census were (*i*) reporting of data separately for rural and urban areas and (*ii*) verification of the reported data by a sample check. For 1961 census the same proforma as in 1956 census was used with the exception in respect of adult males where the sub-head "used for work" was further subdivided into (*i*) Castrated and (*ii*) Uncastrated. Although the changes in the classification of stock introduced for the censuses held after Independence are definite improvements over the one used prior to Independence, it may be pointed out, however, that the sub-division of breeding females into (*i*) In milk and (*ii*) Dry does not seem to be very useful. Because of the seasonality of calvings, the proportion of animals in milk recorded at one point of time in a year cannot be taken as valid for all the seasons and as such has no practical utility. A proper estimate of the number of animals in milk in different seasons can only be obtained through properly designed surveys conducted throughout the year. For the purpose of censuses the classification of breeding females into (*i*) milch stock (those in milk and dry put together) and (*ii*) not calved even once would suffice.

As regards the quality of data collected through livestock censuses the Institute of Agricultural Research Statistics conducted a series of sample surveys to develop a sampling technique for estimating numbers of major categories of livestock, such as, cattle, buffaloes etc., during inter-census years, and providing a scientific and rationalised method of supervision on the work of primary enumeration during census years in order to ensure reliability of the results. It was concluded that the livestock census gives reliable information for broad categories of livestock at the district level (I.C.A.R. Research Series No. 25). According to the Independent sample check on 1956 livestock census conducted by the Directorate of National Sample Survey the percentage differences observed in all India survey figures for cattle and buffaloes over the corresponding census figures varied from 4 to 5 per cent. (Indian Livestock Census (1956), Vol. 1—Summary Tables.)

It is clear from the foregoing that even though the livestock censuses have been conducted almost quinquennially since 1920 ownwards, the coverage of the censuses was not satisfactory until the 1956 census and varied widely from census to census. Besides, owing to territorial redistribution of Provinces by the British prior to Independence and subsequent major reorganizations of the country which took place in 1950 and 1956, many States and districts within States have undergone drastic changes in their geographical content during the period from 1920 to 1961. Perforce the study on trends must be based on the data pertaining to chunks of areas as large as possible which remained undisturbed during the period under reference so as to minimise distortions due to migration. In the region south of Vindhya and Satpura ranges lies the chunk comprising the area covered by the present States of Andhra Pradesh, Madras, Kerala, and 16 districts in each of Mysore and Maharashtra,* which remained almost undisturbed and for which livestock data are available throughout the period. Hence the study on trends could be made with reference to the data obtained for this chunk which is fairly large and can reasonably be considered to be representative of the entire Southern region. However, in the region north of the Vindhya and Satpura ranges, which comprises the remaining part of the country, there is no such large chunk having the characteristics mentioned above. Since practically every State underwent some reorganization after Independence and for none of the States in the North complete data are available, the study has to be confined in this region to the data collected from districts whose geographical content is undisturbed during the period under consideration. The procedure adopted to decide as to what extent a given district has undergone changes in its territorial content, and the limitations of the data collected from such districts as remained almost intact are described in the following sub-section.

2.2. **Collection of Requisite Data.** As direct and authoritative information on the changes which occurred during 1920 to 1961 in the geographical content of the districts in the country was not readily available, the possibility of classification of all the districts with reference to the extent of disturbances in their territorial content by indirect means had to be first examined. This study has been reported elsewhere [Avadhani and Amble (1967)]. The data given in the Table A-II, "Variation in the population during the past fifty/sixty years" of Part II-A Tables of the human population decennial censuses from 1921 to 1961 for each district in the country constitute the main basis of this study. In order to afford a correct basis of comparison, the figures of human population recorded at previous censuses have been adjusted for territorial changes, if any, and tabulated along with the given census count in this Table. So if a district remained undisturbed from 1921 to 1961, then the comparable figures for the district for 1921, '31, '41 and '51, as reported in 1961 census Part II-A Table must necessarily be identical with the corresponding census counts. It is possible to imagine a 'balanced' redistribution of areas between the district in question and its neighbouring districts during the period between two censuses so that the net effect on the total population of the district is zero. A scrutiny of the notes published alongwith the A-II Table on each and every district

---

*Belgaum, *Bijapur* and *Dharwar districts of Mysore*, and Thana, *Dhulia*, Greater Bombay, Kolaba, Poona, Ratnagiri, Satara, Sangli, Kolhapur and Sholapur of Maharashtra could not be taken due to the merger of neighbouring erstwhile Princely States which did not previously participate in the livestock censuses.

showed that such cases did not occur. Hence a district could be taken as totally undisturbed from 1921 to 1961 if the comparable computed population figures for the district (or cluster of districts) given in the A-II Table of the 1961 census Part II-A Tables for 1951, '41, '31 and '21 tallied identically with the respective census counts. Districts (or clusters of districts) which could be taken as undisturbed were put in class A. The districts (or clusters of districts) for which the numerical value of maximum of the relative differences between a census count and the corresponding comparable figure given along with the next census count did not exceed 1 per cent were grouped under the class B, and those for which this difference lay in the range 1-5 per cent under the class C.

It was found that there are only 50 districts which could be considered totally undisturbed and there are 106 districts which are slightly disturbed, but for which the corresponding change in the human population does not exceed 5 per cent in the entire country. Of these 156 districts which belong to the A, B and C classes 108 fall in the North. In order to ensure better representation all the 108 districts belonging to the A, B and C classes are considered for the present study. Of these districts livestock data were not available for 13 districts of which 2 belong to Assam, 1 to Gujarat, 3 to Punjab and 7 to Uttar Pradesh. As a result, Gujarat had to be excluded from the study in question. Further, for all the districts in Bihar and West Bengal for the year 1935, Madhya Pradesh and Punjab for 1925, and Uttar Pradesh for 1940 for which census was not taken, substitutions by linear interpolation between the next later and previous counts were made.

In regard to the collection of information for the selected districts in the North and the chunk in the South data recorded from 1920 to 1961 only in respect of the broad categories of adult males, adult females and young stock of both cattle and buffaloes could be used for the present study since data in more detailed classification were not available for the censuses carried out prior to 1951.

Requisite data secured for all categories from 1920 to 1961 from Agricultural Statistics of India, Vol. 1 and 2, (1920-1940), Indian Agricultural Statistics, Vol. 1 and 2 (1945) and Indian Livestock Census Vol. 2 (1951-61) form the basis of the investigation under consideration.

The methodological aspects of the present study are discussed in the following section.

3. Methodology. Let Y be a growth function of a biological population defined over the time continuum $-\infty < t < +\infty$, and y_{t_1} , y_{t_2} ....y_{t_N} be the valuesof this function observed at different points of time t_i ($i = 1, 2, \ldots\ldots N$). The problem under investigation would amount to finding out the growth curve of Y, that is, the functional relationship between Y and the time t, with the help of the observations y_{t_i}.

To ensure simplicity and ease it is assumed throughout in what follows that the observations y_{t_i} (or y_i for short) are taken at equi-spaced points of time. It may be postulated that when measured on an appropriate scale, each y_i is composed of a trend value Y_i which is the true functional value of Y for $t=i$, a cyclical term C_i with phase, say k, which is characteristic of the environment surrounding the population in question and is independent of the trend, and a random fluctuation E_i with an expectation of zero and independent of both the trend and the cyclical components, so that the additive model, *viz.*,

$$y_i = Y_i + C_i + E_i \qquad \ldots(1)$$

holds good for all $i = 1, 2\ldots, N$.

The simplest and the most effective technique for obtaining the trend values of a time series is the well known device of simple moving averages. When the period of the moving average equals the phase of the cycle, the cyclical terms would be nullified in the moving average, and the random fluctuations would be reduced. If the trend is not linear and k is large, although the error may reduce and cyclical terms may vanish, the resulting moving average centred at i cannot give the ordinate Y_i and it is likely to differ from it considerably, since the simple moving average assumes that the trend values of all y_i's being averaged lie on a straight line. Further the larger the period the greater the loss of number of points for fitting the trend. Hence when the growth curve in a given situation is not expected to be linear, which can be ascertained from the scatter diagram of the observed data, it is desirable to use, specially when no cycle is expected to be present, as small a period as possible for the moving average technique so as to get as close a value to the true trends as feasible, since smaller number of consecutive points can reasonably be assumed to be lying on a straight line segment which is closer to the actual curve. However, it may be mentioned that when cyclical fluctuations are present with phase k in the data, in order to nullify them it is essential to adopt the 'general k-period weighted moving average technique' described by Kendall (1946) for which the basic assumption is that the trend components of k-points being averaged lie on a polynomial curve.

Having obtained the 'approximate' trend values y'_i (say) by the preceding procedure, the next step would be to get the curve which fits well to the poins (i, y'_i), $i = 1,2\ldots\ldots$N. If F(Y, t) is the functional form of the relative growth rate of Y at time t, the growth function Y would be a member of the class of functions defined by the differential equation

$$\frac{1}{Y}\frac{dY}{dt} = F(Y, t) \qquad \ldots(2)$$

For instance, if $F(Y, t) =$ a $(1 - Y/K)$, then substituting in (2) and integrating one obtains the well-known logistic family of growth curves, *viz.*,

$$Y = K/(1 + be^{-at}) \qquad \ldots(3)$$

where b is the constant of integration.

If $F(Y,t) = a f(t)(1 - Y/K)$, where $f(t)$ is a polynomial in t then the growth curve would belong to the Pearl's skew-logistic family (Pearl, *et al.* (1925), which is of the form

$$Y = d + K/[1 + \text{Exp}\ (a_0 + a_1\ t + a_2 t^2 + \ldots\ldots)]. \qquad \ldots(4)$$

If $F(Y,t) = f(t)$, a polynomial of degree m, then the growth function Y belongs to the gencral exponential family,

$$Y = \text{Exp}\ f_1\ (t) \qquad \ldots(5)$$

where $f_1\ (t)$ is the integral of $f(t)$ and consequently a polynomial of degree $m+1$. The functional relationship between log Y and the time t would be a polynomial if and only if the relative growth rate is a polynomial. Also Y would be a linear function in t if the relative growth rate equals $1/t$ at every point of time, or is proportional to $1/Y$.

It is clear that in order to get the functional form of Y with the help of the points (i, Y'_i), $i = 1, 2, \ldots\ldots N$, the simplest procedure would be to try for a suitable polynomial trend in t. If this fails, then a safer alternative in general wolud be to examine the relationship between the relative growth rate, and Y and t by using an approximation such as

$$\left(\frac{1}{Y}\ \frac{dY}{dt}\right)_{t=t_i} = \frac{1}{2}\left[\frac{Y_{t_{i+1}} - Y_{t_i}}{Y_{t_i}\,(t_{i+1} - t_i)} + \frac{Y_{t_i} - Y_{t_{i-1}}}{Y_{t_i}\,(t_i - t_{i-1})}\right] \qquad \ldots(6)$$

which is valid when the time periods t_i are close and then after determining this function and substituting in (2), obtain the required growth curve by integrating the resulting equation.

4. **Trends in the Bovine Population.** In order to facilitate application of the methodology suggested in the preceding section the livestock data collected as described in Section 2 were assumed to refer to equispaced points of time beginning from 1921. The results are not likely to be seriously affected by this assumption.

As the data consist of only 9 observations for each category no meaningful rigorous study could be made to examine the existence or otherwise of cyclical movements in the series. A direct examination of the data for the different categories did not indicate the presence of any cyclical fluctuations. In the light of this to ensure minimum loss of points on the graph, which is essential for the present study, the simple 3-period moving average technique was employed to get the approximations to the true trend components.

In all cases simple polynomial curves such as linear, quadratic, etc., were tried. For the South, trends were fitted to the data on different categories of bovines pertaining to the big chunk to which reference has already been made in Section 2. For the North, to begin with polynomial curves were fitted to the 3-period moving averages of the selected district totals in each State. However, in most of the cases no simple curve was found to explain more than 50 per cent of variation, which could possibly be attributed to differential rate of the inter-state movement of livestock. Therefore, for any category the selected districts totals in each State were first weighted by the ratio of the 1961 livestock census population total for the category in the State and the corresponding total for selected districts and then, after pooling the figures thus obtained for all the States, the exercise was repeated. The percentage variation explained by various curves thus fitted to the data on various categories of cattle and buffaloes in the southern and the northern regions are given in Columns (4) and (5) of Table I. It may be seen from Column (4) of this table that in the southern region the linear curve accounted for only about 39 per cent of variation in the adult males of cattle, but explained more than 90 per cent in the other two categories of cattle. For adult males of buffaloes the linear trend explained about 8 per cent and the quadratic curve about 88 per cent of variation; for adult females 98 per cent of the variation was accounted by the linear curve. But in the case of young stock of buffaloes even the quadratic curve accounted for only 69 per cent of variation. As regards the northern region, it is seen from column (5) of Table I that in the case of all categories of cattle and of adult females of buffaloes it was the quadratic curve which explained about 90 per cent of variation. The picture was the same otherwise as for the southern region.

For describing the trend in a particular category the simplest curve which accounted about 90 per cent of variation was taken as satisfactory. In this sense the trend in adult male cattle in the South was quadratic and for the other two categories of cattle it was linear. Similarly, the trends in adult males, adult females and young stock of buffaloes in the South were quadratic, linear and cubic respectively. The equations of the growth curves thus chosen for all categories of bovines both in the South as well as in the North are presented in Column (3) and Column (4) of Table II.

A question may be raised as to whether the procedure adopted in the case of the northern region for determining the trends of various categories of bovines is at all satisfactory in the light of the fact that the data pertain to groups of districts and not to a large contiguous chunk as in the South and consequently migration effects would not necessarily be eliminated. To examine this to the limited extent possible from available data a similar method was adopted to pool the data pertaining to the selected districts falling in the southern states, *viz.*, Andhra Pradesh, Kerala, Madras, Maharashtra and Mysore, and then the exercise repeated with reference to the pooled data so obtained. The percentage variation accounted by different curves fitted to data on various categories of bovines presented in Col. (6) of Table I and the equations of the appropriate trends in Col. (5) of Table II. A comparison of these results with those obtained in respect of the 'chunk' selected from the southern region shows that in all categories the same curves with almost the same equations describe the trends in both cases except in the case of the young stock of both cattle and buffaloes, where although the curves are of the same nature, the corresponding equations differ slightly.

The growth curves fitted for all categories of bovines both in the South and in the North are depicted in Charts I and II. Of the trends fitted three, *viz.*, those for adult females and young-stock in cattle and adult female buffaloes, all for the southern region, turned out to be linear. According to these trends the population increased quinquennially at the rate of 0·73 million in adult cows, 0·29 million in adult she-buffaloes and 0·57 million in young stock of cattle. In the case of adult males of both cattle and buffaloes in the South and all categories of cattle in the North there was a dip around 1935, while for adult males and females of buffaloes in the North the minimum value of the trend was reached around 1930 and a little earlier than 1925, respectively. The young stock population of buffaloes in both the sectors appeared to be almost stationary during the decade ending around 1950 and thereafter it increased.

In spite of sustained efforts it was not possible to secure any plausible explanation by way of underlying causes of these features of growth curves. It is suggested that the State Animal Husbandry Departments should call for regular seasonal reports from the District Animal Husbandry Officers giving particulars regarding the conditions of season, grazing availability, any extraordinary outbreaks of diseases, etc. and prepare a summary report for the State as a whole based on these. Such reports will be extremely useful in interpreting the livestock census data.

5. **Applications.** The inter-censal livestock numbers, which are necessary for working out the contribution of the animal husbandry sector to national income can be easily worked out from these curves by interpolation. They can also be used for working out projections of livestobk numbers at future periods under the assumption that the relative growth rates which govern these curves would persist then also.

As an illustration projections of bovines have been computed for the years 1966, 1969, 1974, and are given in Table III. It is seen from the table that the projected figures for buffaloe young stock would be as high as 37·8 million by 1974, which appears to be untenable with slow growth in the adult females. This anamoly perhaps might be due to predominantly large effects of non-sampling errors on the young stock figures prior to Independence as has been pointed out in the sub-section 2.1. While searching for an alternative it was observed that the proportion of young stock of both cattle and buffaloes to adult females did not show any particular trend over the years. The figures for young stock were accordingly projected on the assumption that the proportion of young stock to adult females would remain upto 1974 as in 1956 and 1961. The projected figures of young stock of both cattle and buffaloes thus obtained were as follows :

Young stock (million)

Year	*Cattle*	*Buffaloes*
1966	52·5	18·6
1969	55·0	19·1
1975	59·6	20·3

It will be seen from a comparison of these figures with those in table III that this set of projections for cattle do not differ much from the earlier ones based on trends.

Taking these together with the projections worked out in respect of adult bovines from the trends fitted the total bovine population by 1974 would be 274·6 million comprising 216.3 million cattle and 58·3 million buffaloes.

As an alternative to the procedure adopted in what precedes one could use the latest two quinquennial livestock census data for 1956 and 1961 for working out the projections as has been done by Panse and Amble (1965), since these censuses were well conducted and had nearly complete coverage. Assuming that the growth rates in different categories of bovines into which the data were classified in 1956 and 1961 censues would remain the same upto 1974 as in 1956-1961, the livestock population was first projected separately for rural and urban areas and then combined to get the projected figures for the country. The projections thus obtained are shown in Table IV. To facilitate comparison these projections for cattle and buffaloes and those obtained on the basis of the trends are presented in what follows :

Table showing the two sets of projections

	Cattle		*Buffaloes*		*Total Bovines*	
Year	*A*	*B*	*A*	*B*	*A*	*B*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1961*	175·8		51·2		226·5	
1966	189·2	195·0	52·5	58·5	241·7	253·5
1969	198·7	203·3	54·4	63·7	253·1	272·0
1974	216·3	232·4	58·3	73·1	274 6	305·5

A—Projection based on the trends

B—Projection based on the 1961 to 1956 ratio

*—Actual census count

It is seen from the above table that the projections based on the ratio of 1961 population to that of 1956 are consistently higher than the corresponding figures obtained on the basis of trends. The two sets of projections may perhaps be considered as the limits between which the population figures are likely to lie.

6. Summary: All quinquennial livestock census data available until 1961 were sifted and assembled to determine the trends in various categories of animals among cattle and buffaloes. Data were collected category-wise for regions which are as large as possible and remained undisturbed during the period from 1921 to 1961. In the region South of the Vindhya and Satpura ranges the chunk comprising the area covered by the present States of Andhra Pradesh, Madras, Kerala and 16 districts in each of Mysore and Maharashtra, was found to be undisturbed during the period under reference for which livestock data were available. In the North data could be secured only from isolated districts scattered in the present States of Rajasthan, Punjab, Uttar Pradesh, Bihar, Madhya Pradesh, West Bengal and Assam. The study on trends in various categories of bovines was consequently confined to the South and the North separately. The simple 3-period moving average technique was employed to minimise erratic fluctuations and then polynomial curves were fitted. With appropriate adjustments these trends have been utilised for working out projections of livestock numbers for the years 1966, 1969 and 1974. These have been compared with the corresponding figures obtained on the basis of growth at the same rate as between 1956 and 1961.

Acknowledgments : The computational assistance rendered by Shri Shachindra Sharma and Mrs. B. George is thankfully acknowledged.

REFERENCES

1. Avadhani, M.S. and Amble, V.N. (1967): A Note on the Selection of Districs from Different States for Studies of Trends over time—Communicated.
2. Kendall, M.G. (1946): The Advanced Theory of Statistics, Vol. 2, Charles Griffin and Co., London.
3. Panse, V.G. and Amble, V.N. (1965): The Future of India's Population and Food Supply, World Population Conference, Belgrade.
4. Pearl, R. and Reed, L.J. (1925): Skew-growth Curves, Proc. Nat. Acad. Sci. 11.
5. Agricultural Statistics of India (1920-1940), Vols. 1 and 2, Govt. of India Publication.

6. Indian Agricultural Statistics (1945), Vols. 1 and 2, Govt. of India Publication.

7. Indian Livestock Census (1951-1961), Vol. 2, Directorate of Economics and Statistics, Ministry of Food and Agriculture, Govt. of India.

8. Indian Livestock Census (1956), Vol. 1 – Summary Tables, Directorate of Economics and Statistics, Ministry of Food and Agriculture, Govt. of India.

9. Sample Surveys for Improvement of Livestock Statistics (1961): I.C.A.R. Research Series No. 25.

Table I. Percentage Variation Explained by Polynomial Curves.

Sl. No.	*Category*	*Curve*	*Southern Region*	*Selected Districts in*	
				Northern States pooled[2]	*Southern States Pooled*[2]
1	*2*	*3*	*4*	*5*	*6*
	Cattle				
1.	Adult males	Linear	39	64	43
		Quadratic	94	92	97
2.	Adult females	Linear	96	24	98
		Quadratic	—	88	—
3.	Youngstock	Linear	92	61	95
		Quadratic	—	93	—
	Buffaloes				
4.	Adult males	Linear	8	77	6
		Quadratic	88	96	88
5.	Adult-females	Linear	98	85	97
		Quadratic	—	93	—
6.	Youngstock	Linear	57	68	43
		Quadratic	69	69	60
		Cubic	92	94	99

1. Pertains to the chunk selected from the region south of the Vindhya and Satpura ranges (vide section 2·1).

2. For details vide Section 2·1.

Table II. Equations of the Trends Fitted for Different categories of Bovines

Sl. No.	*Category*	*Southern Region*	*Northern States Pooled*	*Southern States Pooled*
1	*2*	*3*	*4*	*5*
	Cattle			
1.	Adult males	$Y=23\cdot2661-1\cdot4205t+0\cdot2166t^2$	$Y=38\cdot8517-1\cdot8051t+0\cdot3336t^2$	$Y=23\cdot2483-1\cdot1819t+0\cdot2253t^2$
2.	Adult females	$Y=14\cdot1441+0\cdot7307t$	$Y=31\cdot9798-2\cdot3390t+0\cdot3371t^2$	$Y=14\cdot0803+0\cdot7854t$
3.	Youngstock	$Y=11\cdot0446+0\cdot5659t$	$Y=28\cdot3677-1\cdot3607t+0\cdot2439t^2$	$Y=10\cdot6984+0\cdot7339t$
	Buffaloes			
4.	Adult males	$Y=2\cdot6568-0\cdot2545t+0\cdot0341t^2$	$Y=3\cdot8547-0\cdot1808t+0\cdot0397t^2$	$Y=2\cdot8339-0\cdot342\ t+0\cdot0414t^2$
5.	Adult females	$Y=5\cdot5496+0\cdot2917t$	$Y=12\cdot8832-0\cdot1654t+0\cdot0716t^2$	$Y=6\cdot1039+0\cdot2352t$
6.	Youngstock	$Y=3\cdot2460+1\cdot2966t-0\cdot3063t^2+0\cdot0231t^3$	$Y=7\cdot3903+2\cdot8200t-0\cdot7463t^2+0\cdot0624t^3$	$Y=2\cdot4235+1\cdot8101t-0\cdot4509t^2+0\ 0345t^3$

Scale. Y in millions and t in 5 year units with $t=0$ at 1921.

Table III. Livestock projections as Indicated by Trends Fitted to Various Categories of Bovines (million)

Year	Cattle			Buffaloes		
	Adult males	*Adult females*	*Young stock*	*Adult males*	*Adult females*	*Young stock*
1	*2*	*3*	*4*	*5*	*6*	*7*
1961*	72·5	54·2	48·8	7·7	25·0	18·5
1966	77·7	59·0	52·0	8·6	25·4	24·7
1969	81·9	61·7	54·3	9·1	26·2	28·7
1974	89·7	67·0	58·4	10·2	27·8	37·8

*Actual census count.

Table IV. Livestock projections based on 1956 and 1961 census data only (millions)

Year	Cattle			Buffaloes		
	Adult males	*Adult females*	*Youngstock*	*Adult males*	*Adult females*	*Youngstock*
1			*4*	*5*	*6*	*7*
1961*	72·5	54·2	48·8	7·7	25 0	18·5
1966	81·4	59·1	54·5	9·1	28·1	21·3
1969	87·6	62·4	58·3	10·1	30·2	23·4
1974	98·7	68·5	65·2	12·0	34·0	27·1

*Actual census count.

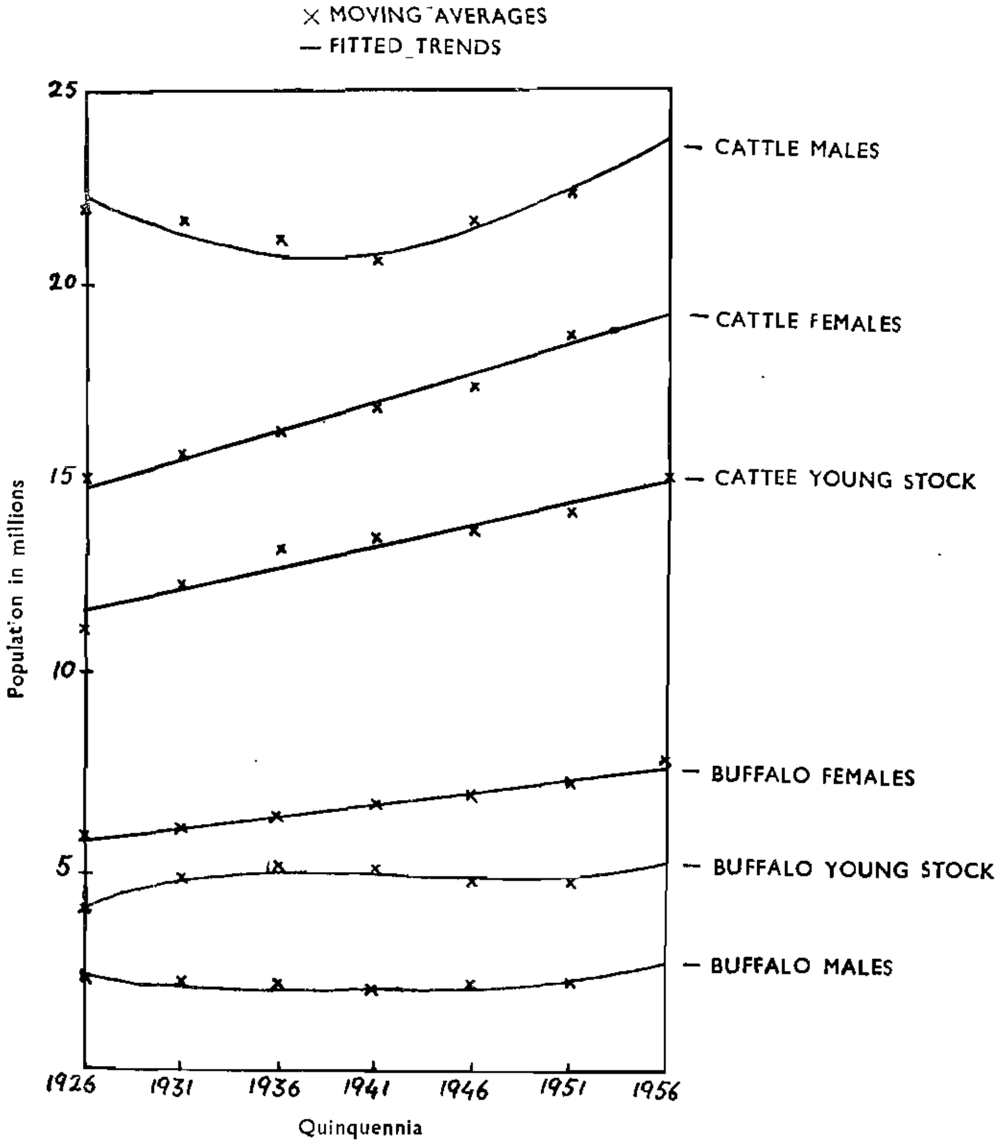


CHART I : Trends in the bovine populations pertaining to the southern region.

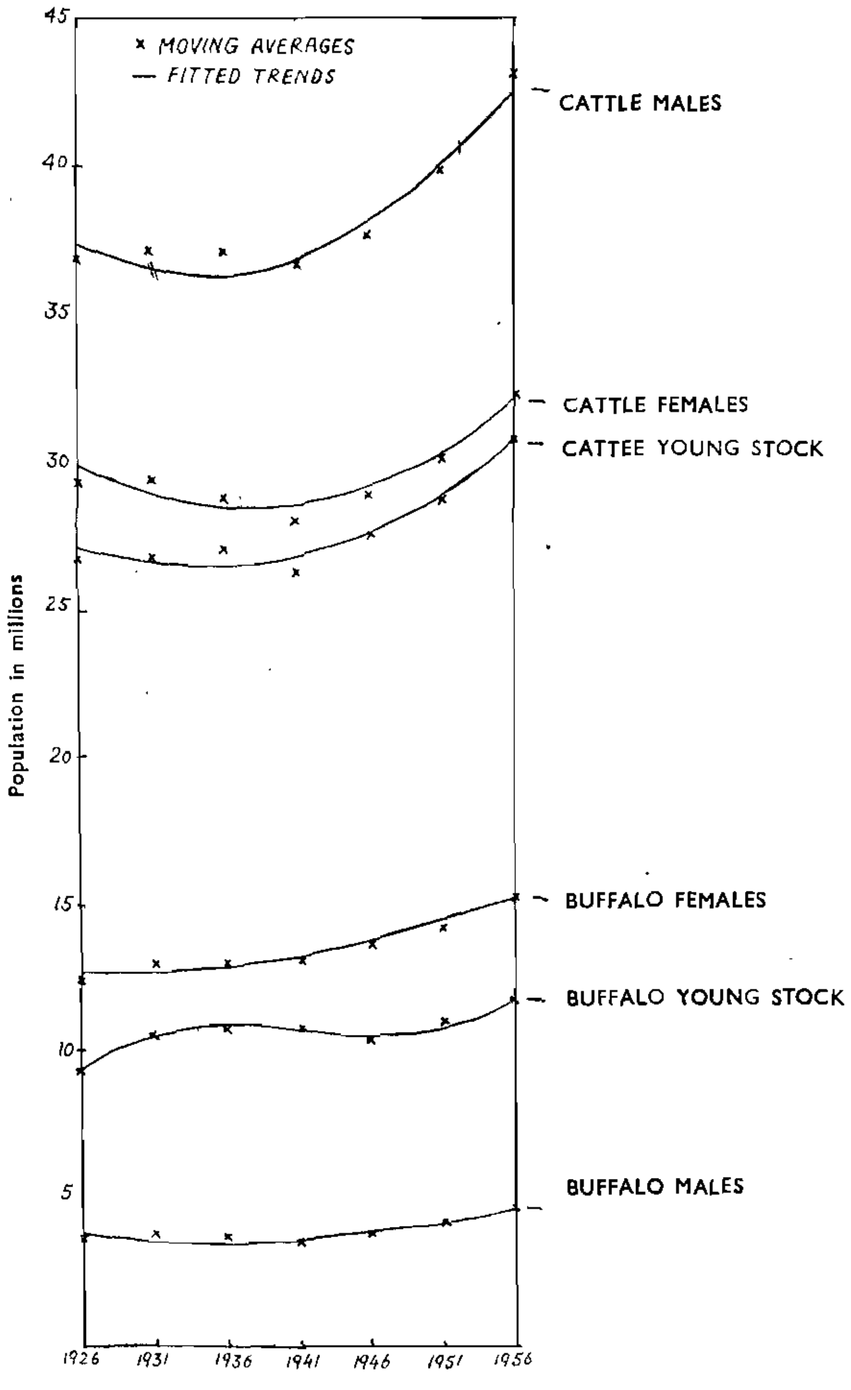


CHART II : Trends in the bovine populations pertaining to the northern region.

APPROXIMATIONS TO THE HYPERGEOMETRIC DISTRIBUTION IN ACCEPTANCE SAMPLING PROCEDURES

By

P.K. BOSE AND S.P. MUKHERJEE,

CALCUTTA UNIVERSITY

1. Introduction. Acceptance sampling plans are based on the distribution of the number of defective items observed in a random sample of n items drawn without replacement from a lot containing N items, a fraction p whereof is defective. The probability of x defective is given by the hypergeometric probability function

$$h(x\ ;\ N, p, n) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}};\ q=1-p,\ x=0,\ 1,\ 2,\ldots\ldots n \quad \ldots(1)$$

Obviously, $\sum\limits_{x=0}^{n} h\ (x\ ;\ N\ ;\ p,\ n)=1$ and $h\ (x\ ;\ N,\ p,\ n)=0$ whenever $x>Np$.

If we concentrate on large lots $(N>>n)$ only, the distribution in (1) can be reasonably approximated by the binomial probability function

$$b(x\ ;\ n\ p)=\binom{n}{x}\ p^x\ q^{n-x}\ . \quad \ldots(2)$$

When the sample is infinitely large and the lot fraction defective very small, so that the number of defective units in the sample is only moderate, the Poisson probability function

$$p\ (x\ ;\ m)=e^{-m}\ \frac{m^x}{x\ !}$$

provides another approximation to (1).

It has been held (Working, 1943) that the binomial approximation can be satisfactorily used whenever $\frac{n}{N}\leqslant 1$. The adequacy of this approximation does not depend directly on p. However, in case where $p\leqslant\frac{n}{N}=f$ the binomial probability function in (2) gives non-zero probabilities of getting more defectives in the sample than are contained in the lot. Consequently $\sum_{x=0}^{Np} b\,(x\,;\,n,\,p) \leqslant 1$ in such situations and this is definitely undesirable. Table 1 shows the sums $\sum_{0}^{Np} b\,(x\,;\,n,\,p)$ for various n, N and p.

It seems that the f-binomial distribution (Coggins, 1928) defined as

$$f\,(x\,;\,N,\,p,\,n)=\binom{Np}{x}\left(\frac{n}{N}\right)^{x}\left(1-\frac{n}{N}\right)^{Np-x} \qquad \ldots(4)$$

may provide a better approximation in cases where $n<Np$.

Evidently, $f\,(x\,;\,N,\,p,\,n)=0$ if $x \geqslant Np$, and thus

$$\sum_{0}^{Np} f(x\,;\,N,\,p,\,n)=1.$$

The conventional binomial and the f-binomial distributions are thus identical when $p=n/N$.

Cowden (1957) studied the cumulative probability of d defectives or less by hypergeometric, binomial and f-binomial distributions for a few combinations of the parameters n, N and p and suggested that when p is small enough, f-binomial gives a closer approximation to the hypergeometric than the conventional binomial.

The object of the present investigation is to compare the binomial and f-binomial probability functions as approximations to the hypergeometric. Two bases for such a comparison are the relation between moments of lower order for these three distributions and the probability of getting d defectives or less. From both these aspects, the f-binomial is a better approximation to the hypergeometric than the ordinary binomial only for the range

$$\frac{1}{N} \leqslant p \leqslant f.$$

2. Comparison of Moments

Let us consider the first three moments of the three distributions spoken of in the earlier section. Expressions for these in the Hypergeometric distribution are

$$\mu'_1\,(h) = np.$$

$$\mu_2\,(h) = \frac{N-n}{N-1}\,npq \qquad \ldots(5)$$

$$\mu_3\,(h) = \frac{N-2n}{N-2}(q-p) \times \frac{N-n}{N-1}\,npq.$$

The corresponding moments for the conventional binomial distribution are $\mu'_1 (b) = np$

$$\mu_2 (b) = npq$$

$$\mu_3 (b) = npq\ (q-p) \qquad .. (6)$$

and those for the f-binomial, as can be easily verified, are

$$\mu'_1 (f) = np$$

$$\mu_2 (f) = np\left(1-\frac{n}{N}\right)$$

$$\mu_3 (f) = np\left(1-\frac{n}{N}\right)\left(1-\frac{2n}{N}\right) \qquad ...(7)$$

Evidently, means are the same in all the three cases. Regarding dispersion the f-binomial disiribution will be a closer approximation to the hypergeometric than the simple binomial if

$$| \mu_2 (b) - \mu_2 (h) | > | \mu_2 (f) - \mu_2 (h) | .$$

Obviously, $\mu_2 (h) < \mu_2 (b)$ so long as $n > 1$.

Now $\mu_2 (h) < \mu_2 (f)$ if $\dfrac{np\left(1-\frac{n}{N}\right)}{\frac{N-n}{N-1}\ npq} > 1$

or if $\dfrac{N-1}{N} \cdot \dfrac{1}{q} > 1$

i.e., if $q < \dfrac{N-1}{N}$

or if $p > \dfrac{1}{N}$.

Also, $\mu_2 (f) < \mu_2 (b)$ if $q > 1 - \dfrac{n}{N}$ or if $p < \dfrac{n}{N}$.

Therefore, over the range $\dfrac{1}{N} < p < \dfrac{n}{N}$, $\mu_2 (f)$ is closer to $\mu_2 (h)$ than $\mu_2 (b)$, the hypergeometric distribution having the
Considering large lots, the condition for this is, therefore
defective in the lot should neither be very small nor large
fraction f.

Thus the dispersion of the f-binomial distributic
that of the hypergeometric if the fraction defective (p) is le
ing fraction $\left(f = \dfrac{n}{N}\right)$.

ıth
ılbert

Quite obviously $\mu_3(h) < \mu_3(b)$. For $\mu_3(h)$ to be less than $\mu_3(f)$ we require

$$\frac{N-n}{N-1}\cdot\frac{N-2n}{N-2}\cdot npq\,(q-p) < np\left(1-\frac{n}{N}\right)\left(1-\frac{2n}{N}\right)$$

or
$$\frac{q\,(q-p)}{(N-1)(N-2)} < \frac{1}{N^2}$$

or
$$(1-p)\,(1-2p) < \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)$$

or
$$3\left(\frac{1}{N}-p\right) < 2\left(\frac{1}{N^2}-p^2\right)$$

or
$$\left[3-2\left(\frac{1}{N}+p\right)\right] < 0.$$

This means either

$$p > \frac{1}{N} \text{ and } p < \frac{3}{2}-\frac{1}{N} \text{ or } p < \frac{1}{N}$$

and
$$p > \frac{3}{2}-\frac{1}{N}.$$

Since p cannot exceed

$$\frac{3}{2}-\frac{1}{N} \text{ for } N > 2,$$

the hypergeometric distribution has the smallest third moment over the range $p > \frac{1}{N}$.

Now, for $\mu_3(f) < \mu_3(b)$, we should have

$$q(q-p) > \left(1-\frac{n}{N}\right)\left(1-\frac{2n}{N}\right)$$

or
$$\left(p-\frac{n}{N}\right)\left(2p+\frac{2n}{N}-3\right) > 0$$

i.e., either $p > \frac{n}{N}$ and $p > \frac{3}{2}-\frac{n}{N}$

or $p < \frac{n}{N}$ and $p < \frac{3}{2}-\frac{n}{N}$.

These two conditions cannot materialise simultaneously. In fact the condition $p > \frac{3}{2}-\frac{n}{N}$ is hardly realised since for $p < 1$ this would require

$n > \frac{N}{2}$. Obviously if $\frac{1}{N} < p < \frac{n}{N}$ the third moment of the f-binomial distribution is closer to that of hypergeometric distribution.

Regarding skewness, it may be noted from the expressions of the third moment that the hypergeometric and the ordinary binomial distributions are positively skew for $p < \frac{1}{2}$, while for $p > \frac{1}{2}$ there is negative skewness. The f-binomial distribution, on the other hand, is always positively skew assuming $n < \frac{N}{2}$. Unlike the hypergeometric or the ordinary binomial which are symmetrical for $p = \frac{1}{2}$, the f-binomial distribution is asymmetrical for all p, unless $n > N/2$.

3. Comparison of O.C. Functions

In a single sampling inspection plan if C denotes the acceptance number, the operating characteristic function of the plan is given by

$$\beta\,(p) = \sum_{x=0}^{C} h\,(x\,;\,N, p, n) \qquad \ldots(8)$$

For various combinations of N, n and p values of $\beta\,(p)$ have been compared with values $\beta_1\,(p)$ and $\beta_2\,(p)$ obtained by using the binomial and the f-binomial distributions respectively in Table 3. An analysis clearly reveals that if $p < \frac{n}{N}$, cumulative probabilities for the f-binomial are closer to the corresponding quantities of the hypergeometric distribution.

4. Discussion of Tables

Table 1 shows that the difference $1 - \sum_{0}^{Np} b\,(x, n, p)$ increases with n for

TABLE 1

Values of $\left[1 - \sum_{x=0}^{Np} b(x, n, p)\right] \times 10^5$

N	50		100			250			
n	5	10	5	10	20	5	10	25	50
$\frac{n}{N}$ / p	·10	·20	·05	·10	·20	·02	·04	·10	·20
·020	384	1617	98	426	1686		0	1	48
·040	60	621	8	86	707		0	0	0
·060	6	203	0	2	96				0
·080	0	58		0	11				
·10	0	15		0	1				
·20		0			0				

fixed N and p, it, however, decreases with p for fixed n and N and also with N for fixed n and p. As N gets larger, the difference is almost zero for larger and larger values of $p-\frac{n}{N}$.

Table 2 gives values of μ_2 and μ_3 for the three distributions concereed for various d combinations of n, N and p. It is found that for the Binomial distribution, for a fixed ratio $\frac{n}{N}$, larger the sample size n, greater is the difference in moments from those of the Hypergeometric distribution. For fixed n and N (and hence f) errors in μ_2 and μ_3 are practically constant for variations in p. For fixed n, errors are smaller for larger lot size (N). While approximating the Hypergeometric distribution by the f-Binomial distribution, however, errors in μ_2 and μ_3 gradually increases with p. Obviously, for smaller values of p, f-Binomial gives moments closer than those of the ordinary Binomial to those of the Hypergeometric distribution. One striking feature is that errors in μ_2 and μ_3 using the f-Binomial distribution take big leaps after $p=0{\cdot}1$.

An examination of entries in Table 3 presents the interesting feature that for $p < \frac{n}{N}$ the sum $\beta_2(p)$ lies between $\beta_1(p)$ and $\beta\,(p)$. Really in this case $\beta(p) < \beta_2(p) < \beta_1(p)$ for smaller values of C and $\beta(p) > \beta_2(p) > \beta_1(p)$

TABLE 3

Cumulative Probabilities of the Hypergeometric, Binomial and f-Binomial distribution

$N=100,\ n=10,\ \frac{n}{N}={\cdot}1,\ p={\cdot}02,\ N=100,\ n=10,\ \frac{n}{N}={\cdot}1,\ p={\cdot}2\ N=250,\ n=20,\ \frac{n}{N}={\cdot}08,\ p={\cdot}02$
$N=250,\ n=20,\ \frac{n}{N}={\cdot}08,\ p={\cdot}10$

	H	*B*	*f-B*	*H*	*B*	*f-B*	*H*	*B*	*f-B*	*H*	*B*	*f-B*
0	·8091	·8171	·8100	·0951	·1074	·1216	·6568	·6676	·6591	·1112	·1216	·1244
1	·9909	·9839	·9900	·3630	·3758	·3918	·9474	·9401	·9456	·3810	·3918	·3947
2	1·0000	·9992	1·0000	·6811	·6778	·6770	·9960	·9929	·9955	·6783	·6769	·6768
3	1·0000	1·0000	1·0000	·8902	·8791	·8671	·9999	·9994	·9998	·8755	·8671	·8649
4				·9743	·9672	·9569	1·0000	1·0000	1·0000	·9637	·9568	·9549
5				·9958	·9936	·9888	1·0000	1·0000	1·0000	·9919	·9888	·9877
6				·9994	·9991	·9977				·9986	·9976	·9972
7				·9999	·9999	·9996				·9998	·9996	·9995
8				1·0000	1·0000	·9999				1·0000	1 0000	·9999
9				1·0000	1·0000	1·0000				1·0000	1·0000	1·0000
10											1·0000	1·0000

TABLE 2

Percentage relative error in the second and third moments in replacing h(x ; N, p, n) by b(x ; n, p) and f(x ; N, p, n)

	N=50, n=5, n/N=·10				*N=100, n=10, n/N=·10*				*N=100, n=20, n/N=·2*				*N=500, n=10, n/N=·02*			
p	bino.	*f*-bino	bino	*f*-bino	bino	*f*-bino	bino	*f*-bino	bino	*f*-bino	bino	*f*-bino	bino	*f*-bino	bino	*f*-bino
·01	8·79	1·10	30·38	3·23	10·00	0·00	34·72	0·00	23·75	0·00	102·08	0·00	1·85	0·82	5·15	2·39
·02	8·89	0·00	30·69	0·00	8·99	1·01	34·72	3·08	23·75	1·61	106·87	3·11	1·92	1·92	5·79	5·79
·03	8·91	1 05	27·97	1·03	10·02	2·08	34·80	6·46	23 75	2·06	102·14	6·39	1·82	2·87	5·60	8 96
·04	9·15	2·33	31·01	6·83	10·00	3·12	34·74	9·84	23·75	3·13	102·06	9·84	1·83	3·95	5·65	11·93
·05	8·90	3·16	30·69	10·02	10·00	4·21	34·77	13·49	23·75	4·21	104·30	13·48	1·84	5·06	5·66	16·26
·06	8·88	4·25	30·70	13·74	10·01	5·32	34·75	17·30	23·75	5·32	102·20	17·30	1·84	6·18	5·66	20·18
·07	8·90	5·39	30·67	17·65	10·00	6·45	34·79	21·33	23·75	6·45	102·13	21·33	1·83	7·30	5·64	24·26
·08	8·88	6·51	30·64	21·72	10·00	7·61	34·74	25·54	23·75	7·61	102·13	25·54	1·84	8·48	5 64	28·61
·09	8·88	7·68	30·66	26·07	10·01	8·80	34·75	30·22	23·75	8·80	102·14	30·20	1·84	9·67	5·66	33·21
·10	8·88	8·88	30·67	30·67	10·00	10·00	34·76	34·76	23·75	10·00	102·13	34·76	1·83	10·88	5·65	30·05
·15	8·88	16·29	30·65	61·01	10·00	16·47	34·76	63·07	23·75	16·47	102·15	63·07	1·84	17·41	5·65	67·05
·25	8·89	30·66	30·95	50·91	10·00	32·00	34·76	158·73	23·75	32·00	102·16	158·73			5·65	165·04

for larger values of C. In cases where $p > \frac{n}{N}$, $\beta(p) < \beta_1(p) < \beta_2(p)$ or the reverse of it. As easily expected, differences are smaller for larger C and rapidly so. However, for the same limit C, the differences are larger for larger values of p.

5. Use in Sampling Inspection Plans

The foregoing discussion makes it evident that probabilities of getting x defectives for the three probability functions are quite close to one another. Closer still are the cumulative probabilities. However, if a choice has to be effected between the ordinary binomial and the f-binomial as an approximation to the Hypergeometric, the latter has to be preferred for cases where

$$\frac{1}{N} < p < \frac{n}{N}.$$

Numerical values of O.C. and A.S.N. functions using the f-Binomial approximation are nearly the same as those obtainable from the Hypergeometric distribution.

ACKNOWLEDGMENT

The authors acknowledge with thanks the help of their student Shri B. Das in preparing this paper.

REFERENCES

1. Coggins, P.B. (1928) : Bell System Technical Journal, Vol. VII, pp. 26-69.
2. Cowden, D.J. (1960) : Statistical Methods for Quality Control, Asia Publishing House, Bombay.
3. Working, H. (1943) : A Guide to Utilization of the Binomial and Poisson Distributions in Industrial Quality Control, Stanford University Press, p. 11.

PARTIAL DIALLEL CROSSES AND INCOMPLETE BLOCK DESIGNS

By

M. N. DAS AND K. SIVARAM

INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS

NEW DELHI

1. Introduction. In order to understand the genetic architecture and to estimate the combining capacities of highly inbred parental lines, suitable single crosses among them are studied. All possible single crosses among the lines give rise to what is called a complete diallel. With the increase in number of parental lines, the number of crosses increases very rapidly leading to the problems of resources and organizational difficulties. In crops like wheat and linseed where the number of seeds per reproductive unit is very low, complete diallel set with larger number of parents becomes unmanageable. Therefore, the alternatives left are either to limit the number of parents or base the study only on a sample of the full diallel, that is on a partial diallel. While discussing the advantages of using partial diallels among a large number of parents as against making all possible crosses among a smaller selected number of parents, Kempthorne and Curnow (1961) cited the following in favour of the former :

(*i*) the general combining ability of the parents can be estimated more accurately ;

(*ii*) selection can be made among the crosses from a wider range of parents; and

(*iii*) the general combining abilities of a larger number of parents can be estimated. Each parent will be assessed with a relatively low precision but larger genetic gains may result from the more intense selection that can be applied to the parents.

Many incomplete block designs have been made use of in sampling the diallel. An interesting feature is that these designs bear one-to-one correspondence with the diallel 'table' and this enables the selection of balanced or partially balanced samples by using these designs.

Balanced incomplete block (B.I.B.) designs in 2 plot-blocks have been used to obtain complete diallel sets while fractions of the diallel have been obtained through partially balanced incomplete block (P.B.I.B.) designs with two plot-blocks [Curnow (1963), Kempthorne and Curnow (1961), Gilbert (1958), Fife and Gilbert (1963)].

The present investigation aims at providing plans for partial diallel crosses for estimating general combining abilities when reciprocal crosses are assumed identical, by using partially balanced incomplete block designs with any block size (as against block size two already used), any values of λ and any number of associate classes. Expressions giving estimates of general combining abilities (g.c.a.'s) of the participating lines have been provided together with standard errors. It has been seen that a P.B.I.B. design with large number of replications is no draw-back for obtaining plans for partial diallel crosses through them as in the case of agricultural experiments.

2. P.B.I.B. designs in two associate classes and partial diallel crosses with identical reciprocal crosses.

According to Bose and Nair (1939), and later Bose and Shimamoto (1952), a P.B.I.B. design in two associate classes is an arrangement, of v treatments in b blocks, such that :

1. Each of the v treatements occurs r times in the arrangement, which consists of b blocks each of which contains k experimental units. No treatment appears more than once in any block.

2. Every pair among the v treatments occurs together in either λ_1 or λ_2 blocks (and are said to be ith associates, if they occur together in λ_i blocks, $i=1, 2$).

3. There exists a relationship of association between every pair of the v treatments satisfying the following conditions:

(*a*) Any two treatments are either first or second associates.

(*b*) Each treatment has n_1 first and n_2 second associates.

(*c*) Given any two treatments that are ith associates, the number of treatments common to the jth associates of the first and kth associates of the second is p^i_{jk}, and this number is independent of the pair of treatments with which we start. Furthermore, $p^i_{jk}=p^i_{kj}$ $(i, j, k=1, 2)$.

The eight parameters, v, b, r, k, λ_1, λ_2, n_1 and n_2 are known as the primary parameters, and the parameters p^i_{jk} $(i, j, k=1, 2)$ are called the secondary parameters. The secondary parameters may be displayed as elements of two symmetric matrices.

$$P_1=\begin{pmatrix} p_{11}^1 & p_{12}^1 \\ p_{21}^1 & p_{22}^1 \end{pmatrix} \quad \text{and} \quad P_2=\begin{pmatrix} p_{11}^2 & p_{12}^2 \\ p_{21}^2 & p_{22}^2 \end{pmatrix}$$

We shall describe below the use of these designs for obtaining plans for partial diallel crosses.

Let there be v lines under investigation numbered from 1 to v in any order. We shall treat them as the varieties of a P.B.I.B. design. The first associaties (or the second associates, whichever is convenient) of each line (treatment) are written beside the line (treatment) in an ascending order. Any line i ($i=1, 2, \ldots v$) is then crossed with every line j ($j>i$), where j is the first (second) associate of i. Therefore, we get $vn_1/2$ or $vn_2/2$ crosses according as the line i is crossed with its first or second associate lines. We have discussed subsequently the method of estimating the general combining abilities (*g.c.a.*) of the participating lines assuming that the reciprocal crosses are identical.

It may be stressed at this point that the above plans for partial diallels and their analysis are entirely independent of the values of λ's and also of the number of replications in the P.B.I.B. design. The P.B.I.B. designs that are useless for block experiments because of large number of replications can be used with advantage for obtaining plans for partial diallel crosses. A further advantage is achieved from the fact that for most of the numbers of lines P.B.I.B. designs are available and therefore plans for partial diallel crosses are also available for these numbers.

3. Analysis

We shall first brifly deal with the model with which we are concerned in diallel crosses.

Assuming that there is no interest in the selfing of the parental lines themselves and that there are no maternal effects, the mean yield from the crosses between the ith and jth lines is expressed as

$$\bar{y}_{ij} = \mu + t_i + t_j + s_{ij} + \bar{e}_{ij} \qquad \ldots(3{\cdot}1)$$

where μ is the average effect, t_i and t_j are the effects due to lines i and j respectively (these effects are usually called the general combining abilities), s_{ij} is the effect due to the non-additivity of the parental effects (which is usually called the specific combining ability) and $\bar{e}_{ij}$ is the random error which may include error due to plot deviation and also due to segregation within the cross. We shall assume that t_i, s_{ij} and $\bar{e}_{ij}$ are independently normally distributed with zero means and variances σ_t^2, σ_s^2 and σ_e^2.

On the above model, the normal equations for estimating the general combining ability t_i of the i th line are shown below, assuming that the crosses are between the first associates.

$$n_1\mu + n_1 t_i + S_1(t_i) = T_i \qquad (i=1, 2, \ldots\ldots, v) \qquad \ldots(3{\cdot}2)$$

where $S_1(t_i)$ denotes the sum of the g.c.a.'s of lines which are the first associates of i and with which the ith line is crossed, and T_i denotes the total yield of all the crosses with the ith line.

Adding such equations for all the lines with which ith line is crossed, we get

$$n_1^2\mu + n_1 S_1(t_i) + n_1 t_i + p_{11}^1 S_1(t_i) + p_{11}^2 S_2(t_i) = S_1(T_i) \qquad \ldots(3{\cdot}3)$$

where $S_2\ (t_i)$ is the sum of g.c.a.'s of lines not crossed with the i th line and $S_1(T_i)$ is the sum of the totals T_i's of those lines with which the ith line is crossed.

Assuming $\sum_{i=1}^{v} t_i = 0$, we get $\mu = \frac{2G}{vn_1}$ where G is the grand total of the yields of all the crosses, and the equation (3·3) reduces to

$$n_1^2\mu + (n_1 - p_{11}^2)t_i + (n_1 + p_{11}^1 - p_{11}^2)S_1(t_i) = S_1(T_i) \qquad \ldots(3{\cdot}4)$$

Solving (3·2) and (3·4), we get

$$\hat{t}_i = \frac{(n_1 + p_{11}^1 - p_{11}^2)T_i - S_1(T_i) - 2(p_{11}^1 - p_{11}^2)G/v}{(n_1 - 1)(n_1 - p_{11}^2) + n_1 p_{11}^1} \qquad \ldots(3{\cdot}5)$$

The sum of squares due to the g.c.a.'s of the lines is

$$\sum_{i=1}^{v} \hat{t}_i T_i.$$

The variance of the difference between g.c.a.'s of two lines

(*i*) which are not crossed is $\frac{2(n_1 + p_{11}^1 - p_{11}^2)\sigma^2}{(n_1 - 1)(n_1 - p_{11}^2) + n_1 p_{11}^1}$

(*ii*) which are crossed is $\frac{2(n_1 + p_{11}^1 - p_{11}^2 + 1)\sigma^2}{(n_1 - 1)(n_1 - p_{11}^2) + n_1 p_{11}^1}$

where σ^2 is the error variance.

Given any line the first expression above is used to compare its difference from each of the n_2 lines and the second, from each of the n_1 lines. Thus we can get the weighted average of the above two variances taking n_2 and n_1 as weights, and this average is given below.

$$\text{Average variance} = \frac{2[(v-1)(n_1 + p_{11}^1 - p_{11}^2) + n_1]\sigma^2}{(v-1)[(n_1 - 1)(n_1 - p_{11}^2) + n_1 p_{11}^1]} \qquad \ldots(3{\cdot}6)$$

The results of second associate crosses are obtainable from the above results by replacing n_1 by n_2, p_{11}^1 by p_{22}^2 and p_{11}^2 by p_{22}^1.

Example. Consider a P.B.I.B. design in 9 treatments with two associate classes.

1	2	3
1	6	4
1	7	5
6	8	3
6	9	5
7	8	4
7	9	3
2	8	5
2	9	4

The parameters of the above design are

$$v=b=9 \qquad r=k=3$$

$$\lambda_1=1,\ \lambda_2=0 \qquad n_1=6,\ n_2=2$$

$$P_1=\begin{pmatrix}3 & 2\\ & 0\end{pmatrix} \quad \text{and} \quad P_2=\begin{pmatrix}6 & 0\\ & 1\end{pmatrix}$$

When the crosses are between first associate lines we write down the first associates of each line in ascending order as follows :

1— (2, 3, 4, 5, 6, 7)
2— (1, 3, 4, 5, 8, 9)
3— (1, 2, 6, 7, 8, 9)
... ...
... ...

Then crossing line i with every line j ($j>i$) where j is the first associate of i, we obtain the following crosses :

(1 × 2)
(1 × 3)
(1 × 4)
(1 × 5)
(1 × 6)
(1 × 7)
(2 × 3)
(2 × 4)
(2 × 5)
(2 × 8)
(2 × 9)
(3 × 6)
(3 × 7)
(3 × 8)
(3 × 9)
(4 × 6)
(4 × 7)
(4 × 8)
(4 × 9)
(5 × 6)
(5 × 7)
(5 × 8)
(5 × 9)
(6 × 8)
(6 × 9)
(7 × 8)
(7 × 9)

Using the equation (3·5) the solution of t_1 is given by

$$\hat{t}_1=\frac{3T_1-S_1(T_1)+2G/3}{18}$$

Also $V(\hat{t}_1-\hat{t}_2)=\frac{4}{9}\sigma^2$

and $V(\hat{t}_1-\hat{t}_3)=\frac{1}{3}\sigma^2$

The average variance of the difference between any two lines is $5\sigma^2/12$.

While choosing a sample from a diallel it is better to have a connected sample (Clatworthy 1955). If the plans be disconnected, the lines will fall into sets so that no two lines from different sets are crossed. But this need not be a serious drawback as in the case of estimation of treatments through incomplete block designs, because the estimation of g.c.a. is still possible through such disconnected samples by following the method discussed subsequently by Curnow (1963).

4. Use of P.B.I.B. designs with m-associate classes

In previous section we discussed the method of obtaining plans for partial diallel through P.B.I.B. designs in two associate classes. The sample was so formed that each of the lines was crossed with those appearing in one of the two associate classes. This concept may be generalised by using P.B.I.B. designs with m-associate classes. In this case, each line is crossed with each of the lines present in r of the m associate classes.

Let us consider a P.B.I.B. design with v treatments (lines nnmbered in some order) in m-associate classes. Let n_j $(j=1, 2, \ldots, m)$ denote the number of lines present in the jth associate class. For each line i $(i=1, 2, \ldots v)$ a given r associate classes among its m classes are choosen and all the lines in them are pooled. The line i is then crossed with every line j in the pool such that $(j>i)$, Thus we will have a sample of size $v\,(\sum\limits_{(j)} n_j)/2$ where $\sum\limits_{(j)}$ implies summing over the selected r associate classes. For the analysis we shall assume without loss of generality, that each line is crossed with its first r $(r<m)$ associate lines. Let $\sum\limits_{(j)} n_j=N$. Now the normal equations for estimating the *g.c.a.* of the lines through the least squares technique taking the usual model come out as below :

$$N\mu+Nt_i+\sum_{(j)} S_j(t_i)=\sum_{(j)} T_j=Q_i$$

$$(i=1, 2, \ldots, v) \qquad \text{.. (4·1)}$$

where S_j (t_i) is the sum of the g.c.a.'s of lines which are jth associates of the ith line and T_j is the total yield of the jth associate crosses involving the ith line. Adding such equations over the first associate lines of i, we get

$$Nn_i\mu+NS_1(t_i)+\sum_{(j)} p_{1j}{}^1\, S_1(t_i)+\sum_{(j)} p_{1j}{}^2 S_2(t_i)+\ldots$$

$$+\sum p_{1j}{}^m\, S_m(t_i)=S_1(Q_i)$$

where S_j (Q_i) is the sum of the Q's of lines which are jth associate of the ith line.

In general, adding such equations over the kth associates of i, we have.

$$Nn_k\mu+NS_k(t_i)+\sum_{(j)} p_{kj}{}^1\, S_1(t_i)+\sum_{(j)} p_{kj}{}^2\, S_2(t_i)+\ldots\ldots$$

$$+\sum_{(j)} p_{kj}{}^m\, S_m\,(t_i)=S_k(Q_i)$$

$$k=1,\ 2,\ \ldots.,\ (m-1) \qquad \text{...(4·2)}$$

Assuming that $\sum\limits_{i=1}^{v} t_i=0$, equations (4·1) and (4·2) can be solved for t_i and the analysis completed in the usual lines.

A particular case : Use of 3-associate designs

When $m=3$ and $r=1$, we arrive at the simple case of generating plans for partial diallel crosses from a 3 associate P.B.I.B. design by crossing any line with its first associate lines. It has been seen that n_j must be greater than 2 ($j=1, 2, 3$) in order to make the sample of crosses connected. The normal equations for estimating the general combining ability of the lines comes out as below from (4·1) and (4·2).

$$n_1\mu+n_1t_i+S_1(t_i)=T_i \qquad (i=1, 2, \ldots, v) \qquad \ldots(4{\cdot}3)$$

$$n_1^2\mu+n_1t_i+n_1S_1(t_i)+p_{1_1}^{\ 1}\ S_1(t_i)+p_{11}^{\ 2}\ S_2(t_i)+p_{11}^{\ 3}\ S_3(t_i)=S_1(T_i) \qquad \ldots(4{\cdot}4)$$

and $$n_1n_2\mu+n_1S_1(t_i)+p_{12}^{\ 1}\ S_1\ (t_i)+p_{12}^{\ 2}\ S_2\ (t_i)+p_{12}^{\ 3}\ S_3\ (t_i)=S_2(T_i) \qquad ..\ (4{\cdot}5)$$

Assuming $\sum\limits_{i=1}^{v} t_i=0$, we get $\hat{\mu}=2G/vn_1$ and from equations (4·3), (4·4) we get

$$\hat{t}_i=\frac{(A_2B_3-A_3B_2)T_i-B_3\mathrm{Si}(T_i)+A_3S_2(T_i)-\dfrac{2G}{v}(A_2B_3-A_3B_2-n_1B_3+n_1A_3)}{\triangle} \qquad \ldots(4{\cdot}6)$$

where

$$A_1=(n_1-p_{11}^{\ 3}) \qquad B_1=-p_{12}^{\ 3}$$

$$A_2=(n_1+p_{11}^{\ 1}-p_{11}^{\ 3}) \qquad B_2=(p_{12}^{\ 4}-p_{12}^{\ 3})$$

$$A_3=(p_{11}^{\ 2}-p_{11}^{\ 3}) \qquad B_3=(n_1+p_{12}^{\ 2}-p_{12}^{\ 3})$$

and $$\triangle=n_1(A_2B_3-A_3B_2)-(A_1B_3-A_3B_1).$$

The sum of squares due to the g.c.a.'s of the lines is

$$\sum_{i=1}^{v}\hat{t}_iT_i.$$

As $\sum\limits_{i=1}^{v} t_i=0$, no correction factor need be subtracted.

The variance of the difference between g.c.a.'s of two lines is now given by,

$$(i)\ \ V(\hat{t}_i-\hat{t}_{i'})=\frac{2\sigma^2}{\triangle}\ (A_2B_3-A_3B_2+B_3)$$

when line i' is crossed with the ith line

$$(ii)\ \ V(\hat{t}_i-\hat{t}_{i'})=\frac{2\sigma^2}{\triangle}\ (A_2B_3-A_3B_2+A_3)$$

when line i' is not crossed with the ith line but is crossed to a line to which line i is crossed.

$$(iii)\ V(\hat{t}_i - \hat{t}_{i'}) = \frac{2\sigma^2}{\Delta}\,(A_2B_3 - A_3B_2)$$

when line i' is neither crossed with the ith line nor with a line to which i is crossed.

Denoting the above three variances by V_1, V_2 and V_3 respectively, the average variance of the difference between g.c.a.'s of any two lines comes out as below :

$$\text{Average variance} = \frac{n_1V_1 + n_2V_2 + n_3V_3}{n_1 + n_2 + n_3}.$$

It appears that for plans obtainable through the three association designs ($m = 3$, $r = 1$), connectedness *in the samples* is ensured when the following parametric relations are satisfied. When the ith associate lines are crossed then one of the two conditions should be satisfied :

(*i*) Either $p_{ij}^{i} = 0$ for all $j \neq i (j = 1, 2, 3)$

(*ii*) Or, if $p_{ik}^{i}, = 0$ and $p_{im}^{i} > 0$, then

p_{ik}^{m} must be greater than zero.

It may, however, be pointed out that the method of analysis presented here holds even for 'disconnected' samples.

Through 3 associate P.B.I.B. designs we can reduce the total number of crosses when there is a large number of lines. This is so because the total number of lines is $\frac{1}{2}\, vn_i$ where n_i is the number of treatments in the ith associate class. Now if there be two classes the value of n_i is likely to be large. But with the same v if there is a design with three associate classes the values of n_i's are likely to be small and there is more flexibility in the choice of the associate classes.

An Example :

For illustration we have used the two-associate cyclic P.B.I.B. design

$$v = 17,\ b = 34,\ r = 8,\ k = 4,\ n_1 = 8,\ n_2 = 8,\ \lambda_1 = 1,\ \lambda_2 = 2$$

$$P_1 = \begin{bmatrix} 3 & 4 \\ & 4 \end{bmatrix} \quad \text{and} \quad P_2 = \begin{bmatrix} 4 & 4 \\ & 3 \end{bmatrix}$$

as given by Bose et al (1954) and obtained a sample of the diallel involving 17 lines of bajra (Pennisetum typhoides). The character under study is productivity and is measured by giving scores ranging from 3 to 10 with reference to the yield of C.M.S. 24*A* as the standard with a score of 5. The observations which were collected by Shri G. Harinarayana, a Ph.D. student in the Genetics division of I.A.R.I., New Delhi, are presented in Appendix I along with the association scheme of the P.B.I.B. design. The association scheme has the property that the first associates of a treatment i are obtained by adding $(i-1)$ to each of the first associates of treatment 1. Hence it is sufficient to indicate the first associates of the treatment numbered 1. The total scores from crosses involving each of the lines are shown below in Table 1. The values of $S_1(T_i)$ for the different lines are also shown in the same Table together with their g.c.a.

TABLE 1

Lines	*Total Scores* (T_i)	$S_1(T_i)$	*g.c.a.* $\hat{t}_i=[119T_i-17S_1(T_i)+2G]/884$
1	26·0	229·0	−3·5408
2	27·5	228·5	−1·4254
3	31·5	226·0	+4·4401
4	33·0	228·0	+6·0747
5	29·0	227·5	+0·7862
6	25·5	223·5	−3·1562
7	31·0	224·5	+4·0555
8	30·0	225·0	+2·6131
9	27·5	230·0	−1·7138
10	30·0	242·0	−0·6561
11	26·0	229·0	−3·5400
12	25·0	224·5	−4·0215
13	28·5	232·5	−0·8484
14	26·5	234·5	−3·9254
15	30·0	232·0	+1·2670
16	31·0	227·0	+3·5747
17	28·0	224·5	+0·0170

Sum of squares due to g.c.a. is now given by $\sum_{i=1}^{17} \hat{t}_i T_i=11{\cdot}7187$

Analysis of Variance Table

Source	*d.f.*	*S.S.*	*M.S.*	*F(5%)*
g.c.a.	16	11·7187	0·7324	0·6033
Error	51	61·9137	1·2139	
Total	67	73·6324		

The g.c.a. effects do not differ significantly at 5 per cent level of significance.

APPENDIX I

Productivity Scores (Pennisetum Typhoides)

The observations are averaged over two replications

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}
P_1																	
P_2																	
P_3																	
P_4	3·0																
P_5		2·5															
P_6	3·0		3·0														
P_7	4·5	4·0		5·5													
P_8	1·5	3·0	5·5		4·5												
P_9		4·5	2·5	4·0		2·5											
P_{10}			3·0	3·5	4·0		3·5										
P_{11}	3·0			4·0	4·5	3·0		4·0									
P_{12}	3·0	3·5			3·0	3·5	2·5		3·5								
P_{13}	3·0	2·0	4·5			4·5	4·0	3·0		4·5							
P_{14}		3·5	4·5	2·0			4·0	4·0	2·5		2·0						
P_{15}	5·0		4·0	5·5	4·0			4·5	2·0	2·5		2·5					
P_{16}		4·5		5·5	3·5	1·5			6·0	4·5	2·5		3·0				
P_{17}		4·5		3·0	4·5	2·0				4·5	3·0	3·5		3·0			

Scale for the scores: 1 —Poorest
10—Highest
5 —Standard (CMS 24 A)

Association scheme for the P.B.I.B. design in two associates for $v=17$, $b=34$, $n_1=n_2=8$. The first associates of treatment number 1 are (4, 6, 7, 8, 11, 12, 13, 15).

REFERENCES

1. Bose, R.C. and Nair, K.R. (1939) : Partially balanced incomplete block designs. Sankhya 4, 337.
2. Bose, R.C. and Shimamoto, T. (1952) : Classification and analysis of P.B. I.B. designs with 2 associate classes. Jour. Am. Stat. Ass. 47, 151-184.
3. Bose, R.C. Clatworthy W.H., Shrikhande S.S. (1954) : Tables of P.B. designs with two associate classes. Carolina, North, Agric. Expt. Station, Tech. Bull. No. 107.
4. Clatworthy, W.H. (1955) : Partially balanced incomplete block designs with two associate classes ann two treatments per block. Jour. Res. Nat. Bur. Standards 54. 177-190.
5. Curnow, R.N. (1963) : Sampling the diallel crosses. Biometrics 19, 278-306.
6. Fife, J.L. and Gilbert, N.E.G. (1963) : Partial diallel crosses Biometrics 19, 278-286.
7. Kempthorne, O. and Curnow, R.N. (1961) : The partial diallel crosses. Biometrics, 17, 229-250.

THE STATISTICIAN AND PLANNING EXPERIMENTS*

By

D.J. FINNEY**

If you were to ask me to define *statistics*, I should first express reluctance to demarcate rigid limits to a discipline that has proved of immense value to science and technology because its practitioners have been prepared to accept new responsibilities. If pressed, I should assert that statistics is the science of *collecting*, *analysing*, *interpreting* and *presenting* numerical information, under conditions such that unexplained variations between units from which the information is collected are so great as to be important to the interpretation.

My definition is one of perhaps hundreds that have been suggested. I make no claim that it is the best. I am convinced that the four elements—collection, analysis, interpretation, presentation—must be included, and that the science I practise has gained inestimably from their interactions. I am therefore strongly opposed to the fragmentation encouraged by some present trends. Occasionally those whose interests lie mainly in the mathematical theory of statistics imply, a little arrogantly, that they are the elite. Their contributions to the development of new methods of analysis are vital; much that seems at first entirely abstract mathematics eventually becomes part of the corpus of statistical practice. Yet, parhaps the value would be greater if abstractions more often were fertilized by association with real data. Certainly we must avoid attitude that set theoretical studies apart from applied statistics. At present, I am concerned at the growing use of the term "data analysis" as a synonym for applied statistics. If we need to save two syllables, this can serve as an alternative to "statistical analysis". My fear is the encouragement it gives to neglect of the other three elements—collection, interpretation, presentation—with which statisticians must be deeply involved. Once we accept the polarization implied by a distinction between statistics and data analysis, "statistics" will tend to be regarded as the name for a branch of pure mathematics; the vacuum established by the two syllable saving may then be part filled by "*mere* data analysis". In this I see no merits. Certainly the sciences, industries, and technologies that can be helped by statistics will suffer.

I want to say about some aspects of collecting data. Often the statistician must endeavour to make sense of data collected haphazardly, or accumulated without much advance planning. In an important paper nearly three years ago, Cochran (1965) discussed the analysis and inference appropriate when the only available evidence on a problem is of this kind. If

---

*An address to the American Society for Quality Control (Metropolitan Section), 1 December 1967.

**Department of Statistics, University of Edinburgh.

time permits, and other conditions present no insuperable obstacle, the statistician will usually prefer to organize the collection of information by means of planned sampling procedures. He is then able to use standard statistical logic in making inferences about the population from which his sample was drawn. I need not here elaborate on the logic or the importance of sampling. I must emphasize two points. First, the rules adopted for choosing the sample determine on what properties of the population inference is possible. Secondly, however well-designed the sampling plan, the conclusions permissible relate strictly only to the population as it is, and do not include inference about the consequence of change in conditions.

I am here concerned not with general questions of sampling but with the particular needs of experimentation. The fundamental characteristic of an experiment is that the investigator has the power (and the duty) to assign treatments of his choosing to different subjects— animals, machines, patients, or other units for observation and measurement. When he exercises that power in accordance with established principles of experimental design, with proper regard for randomization, valid inference may be drawn about the changes to be expected if one treatment is replaced by another. We sometimes forget the philosophical and practical implications of limiting conclusions to the population from which the subjects were drawn, but these I shall not discuss further here.

During the last half-century, a vast literature of experimental design has been created. This ranges from abstract investigation of combinatorial mathematics to detailed specification of certain simple designs and of the arithmetical steps appropriate to the analysis of results. Far less attention has been given to problems of selecting the design best suited for a particular piece of research, and to even broader questions concerned with embedding good experimental design within a planned strategy for a research project.

In the statistical analysis of a body of records, the statistician must usually be the arbiter of the methods to be adopted. He needs from the investigator information on the data, on how they were collected, and on factors relevant to variation; he must learn enough about the investigation to formulate the right questions; on the choice of methods, however, and on technicalities such as transformations or heterogeneities, he alone is the expert.

Very different is the situation during the planning of research. Efficient deployment of resources and effective study of a complex problem demands close collaboration between investigator and statistician. Each makes his special contribution, but each must have sufficient understanding of the other for dialogue and criticism to be beneficial.

Even at the start of planning an experiment, a statistician experienced in the field may be helpful. The definition of treatments needs to be coordinated with the declared objectives, at which point the logic of statistical analysis must be in mind. In a medical context, Schwarz and Lellouch (1967) have usefully distinguished between explanatory and pragmatic experiments. Are alternative treatments (of men, of animals or in an industrial process) to be compared under conditions equalized in respect of concomitant factors, with a view to analytical discussion of causes? This may be appropriate to academic research on the effects of closely related chemical compounds. Or is each "treatment" to be a complex assembly of factors, without any theoretical balance between specifications of different treatments? This is appropriate to a pragmatic study of alternative therapeutic regimes, or of alternative machines that differ widely in structure, the aim being to discover the best for a certain task. To some extent, factorial design offers a way of reconciling the explanatory and pragmatic approaches; nevertheless, the distinction between research intended for the dissection of a chain of causation and research that provides a basis of decisions should affect planning.

I illustrate by reference to a question commonly put to the statistician: what should be the size of a projected experiment ? His first response will be to ask about resources and requirements. If the primary limitation is on total size, as measured by a number of subjects, plots, or other units, he is likely to advise that all be used. If the demand is that certain differences be estimated with specified precision, and if some knowledge of variances exists, the familiar σ^2/n formula can usually be employed to indicate the replication necessary. Greater problems arise when the experiment must be related to the value placed upon a decision that uses its results. Is a change in a manufacturing process justifiable, in terms of higher quality or increased productivity ? Is a new anti-arthritic drug to be recommended in preference to that in current use ? The statistician must take account of information from pilot trials or other available sources. He and the investigator jointly must assess the cost of an experiment in relation to its size ; here they include not merely direct costs but also losses from delay in reaching a decision, or from patients receiving a drug that is not the best. They must express, on an agreed scale, the benefits to be expected from each possible decision, in relation to the magnitude of the true difference between treatments; their formulation must either cover a forward period representing the useful life of the decision or suitable discount benefits that lie far ahead. The strands can then be combined in estimation of the size of experiment that maximises expected net gain. The idea is simple, the execution less so. Many different formulations are possible and much remains to be studied. Net gain is seldom very sensitive to appreciable variations about the optimal size of experiment. However, arguments from approaches as different as those of Grundy, Healy and Rees (1956) and of Colton (1962) agree that the recommendation should usually be *either* "Base a decision on existing information without further experiment" *or* "Conduct a new experiment large enough to give information far in excess of that already available". The common policy of "Let us re-examine the situation after one more experiment of moderate size" is rarely wise !

So much for the pragmatic approach. The explanatory one might be typified by comparisons between alternative amounts of a reagent or alternative temperatures for a chemical process. Here the size of an experiment may be governed by an arbitrary condition that, unless a specified precision is achieved, the work is not worth doing : perhaps research effort would be better diverted into different channels. Alternatively, the attitude may be that knowledge is simply additive ; whatever the size of an experiment, it will add proportionately to existing information on parameters.

Once the size of a new experiment has been decided (in terms of number of experimental units or total expenditure), questions arise that are too often thought not the concern of the statistician. Certainly the choice of treatments must rest largely with the investigator : he determines what the experiment is about ! But the statistician can contribute advice on the number and spacing of levels of continuous factor, on the relevance of control treatments to the desired form of inference, on the merits of multi-factorial design, on blocking and confounding, and so on. For a single continuous variate, such as temperature, the principles of estimating a simple regression equation are well understood : if linearity is in no doubt, concentrate on two widely spaced levels, if deviations from linearity may be important (a more common situation), equal divisions of effort over three or four levels is desirable. However, if a regression is of known non-linear functional form in which one or more parameters must be estimated, the deployment of resources is likely to require special study by a statistician. The optimal policy will usually depend upon the exact objective specified : to estimate a particular parameter as precisely as possible, to predict the response over a range of temperatures, to estimate the temperature that maximizes responses, or to compromise between several aims.

Division of the experimental units into blocks, balancing of treatments over blocks so as to minimise loss of information, multi-factorial design, confounding and fractional replication—these form another large division of the statistician's responsibility, too large for detailed discussion here. To use them well, he must know what characteristics of the experimental material have major effects on variability. The exact definition of factors for inclusion, and the number of factors call upon statistical skill and experience. Choice of a set of factors that adequately explores a complex situation, insertion of enough factors to ensure that experimental resources are fully utilized, and construction of a suitable confounding system may require considerable ingenuity. For example, I work to broad principles such as "include as many factors as possible", "avoid split-plot design", "have every factor at the same number of levels", but always with the reservation "when practicable" ; these are not absolute rules, but guides that must be modified to meet the exigencies of the particular experimental problem. The fact that modern high speed computers remove the need to worry about the labour of analysis for some complex designs occasionally has the consequences that experimenters and statisticians think symmetries of design no longer important. In reality, certain aspects of symmetry can affect greatly the clarity and specificity of interpretation and the balance of information over factors.

Yet another vital consideration is the choice of measurements to be made on each plot or other unit of an experiment. If an experiment is costly in materials and execution, the investigator may be wise to include measurements other than those that relate to his main interest. For example, the statistician may encourage him to record aspects of growth of on organism other than final weight and height, so that growth curves or interrelations of parts can be studied, or to make initial measurements that will improve precision by way of covariance analysis. In an explanatory experiment, probably each variate deserves separate statistical analysis, as also may some functions of two or more variates that represent characteristics of the plots. In a pragmatic experiment, on the other hand, a multivariate analysis of some type (multiple regression, discriminant, or canonical) may facilitate appropriate conclusions even though the function of variates obtained has no obvious interpretation.

I hope I have indicated sufficiently that planning an experiment involves much more than extracting from a catalogue a design conforming to treatments, factors, and blocks dictated by the investigator, and then supplying a randomisation. I conclude by mentioning two specialised examples of planning that illustrate the many points at which statistical thought should impinge.

The aim of biological assay is to compare two materials whose biological effects are identical, except that one behaves as a dilution of the other. Thus x units of weight (or volume) of one material should act exactly as Rx units of the other, where R is the dilution factor. If application of either material to an animal produces a measured response, y, the expectation of which depends upon x, a regression experiment can be set up for the estimation of R. The two materials must have regression functions identical except for the factor R. The experiment tests each material at enough values of x to give adequate estimates of its regression. Statistical analysis then finds a value for R that most nearly superposes the two functions. Simple though this is in concept, several questions arise:

(*i*) What is the form of the regression function ?

(*ii*) What the is nature of the variation of individual responses about the regression ?

(*iii*) How can an experiment with limited resources estimate the regression parameters in such a way as to give the best estimate of R?

(*iv*) How is the correctness of form of the regression function to be tested ?

(*v*) How is the adequacy of superposition of the two functions to be tested ?

Often the answer to the first two questions is a linear regression on x (or on log x), with constant variance. Perhaps too often this is assumed uncritically in a new situation, when in reality a statistician ought first to be consulted on a pilot study of the regression itself. Do we, even in 1967, know enough about how to design an experiment for determining the nature of an unknown regression ? Unless a simple transformation to a linear regression is available, or the particular function has previously been studied in the same context, the third question may demand a careful algebraic examination. The final questions relate to what I term tests of assay validity. The experimenter and the statistician in consultation must assess the strength of pre-existing evidence on the regression function, and on the belief that the two materals are simply related by the dilution factor R. In the light of this assessment, the allocation of resources optimal for estimating R may need to be modified so as to permit tests of significance on the linearity of regressions, on the parallelism of regression equations, or on other requirements essential to the validity of the estimation process. Almost always an experiment planned to permit such tests is less precise in its estimation of R than is an experiment that ignores all need to test validity. Consequently, judgement (in part subjective, but experienced) is required in deciding how far to sacrifice apparent precision to reassurances that the results is meaningful. Statistical theory then helps to determine a design providing significance tests that are powerful in relation to the most likely types of invalidity.

My second example, totally different, is that of planning the multistage screening of a large number of entities. The aim is to discover the one or several which have maximal values of a characteristic that cannot be measured without experimental error. The entities may be chemical compounds under consideration for some form of therapeutic activity, new varieties of sugarcane grown from independent seedlings, or even applicants for scholarships or jobs. The total amount of testing and measuring is limited by the resources available. Screening consists of a succession of operations:

(1) Expose all entities to an experiment or test situation ;

(2) Make an appropriate measurement on each ;

(3) Select a subset of entities as having the highest values for the measurement recorded, and discard the remainder ;

(This completes one stage)

(4) Repeat the process on the selected subset entities so as to reduce to a subset of these ;

(5) Continue in this fashion over several stages;
(6) Take the finally selected small subset as the output of the screening.

Why proceed in this fashion, rather than complete the whole screening in one experiment and one stage of selection ? To select the best 5 out of 1000, why not use all available resources in a single experiment for the 1000, and then take the 5 with highest means ? The reason is that many of the 1000 will be so much inferior to the best 5 that they can be rejected after only a little testing ; the saving represented by this can be used to provide higher replication on the serious contenders for the final set. Consequently one can see intuitively that a first stage with little or no replication of tests on each entity, a second stage in which each of the entities retained is tested with greater replication, and subsequent stages with steadily increasing replication of a decreasing number of entities, must improve the statistical expectation of quality among those finally selected.

Quite commonly what is essentially a screening process may never be brought to a statistician in its completeness ; instead he is consulted only on individual experiments belonging to single stages. Within any stage, the statistician's concern is with the standard questions of design for an experiment that will compare a stated number of treatments (or entities) as precisely as possible. However, the planning of the whole process raises more interesting and more important statistical questions. The total resources available represent a limitation on replication at various stages, perhaps by simple equality to the summation of replicates over all entities and stages, or perhaps by some more complicated function. Subject to this constraint, the number of entities selected at each stage for retesting in the nest and the replication provided for each entity at each stage can be varied. The total number of stages may also be open to choice, though, if each requires one year, account will need to be taken of the cost of delay in reaching the final selection. The criterion of selection at each stage needs to be defined, especially if the measurements are not to be of exactly the same character for every stage. Consideration must be given to whether selection at a stage is to be based solely on the results of testing in that stage, or whether evidence accumulated in previous stages is also to be used and, if so, how. Even the possibility of differential replication *within a stage*, to an extent determined by the earlier evidence on the surviving entities, needs evaluation.

Provided that the variance of comparisons between entities at any stage can be expressed, at least approximately, as a function of the replication (by a σ^2/n type of rule, but this may be complicated by design alternatives), the problem of optimal choice of numbers, replications, and number of stages can be tackled. The analytical mathematics seems usually to be difficult or wholly intractable, except for rather trivial formulations. Computer simulation can be relatively easily adapted to alternative formulations in respect of distributions, criteria, and various parameters.

Rather surprisingly, the system appears to be fairly robust in respect of some conclusions. For example, for any fixed number of stages, a "symmetrical" plan may be suggested. At each stage, the fraction selected for retention in the following stage is approximately the same, and is determined by the fraction to be selected over the whole process ; also, the resources available for testing are divided equally between the stages. Thus a 3-stage process for eventually selecting 4 entities out of 500 might be performed by successive 20 per cent selections of 500 to 100 to 20 to 4, with 5-fold increase of replication per entity at each stage. Though not the true optimal, this seems commonly to be close enough for practical purposes. Improvement in performance coming from increase in the number of stages falls off rapidly, and the advantage of more than 5 stages is seldom appreciable. Similarly, no great gain seem to accrue from using at each stage the evidence of all previous stages. These are not absolute truths, and models and parameters could be propounded for which optimal conditions are very different. Studies so far made of models that, though oversimplified, approximate to real situations indicate that they are good general guides.

I do not wish to leave the impression that this complex range of planning problems is solved. I hope I leave the reader with my own belief that it includes questions of legitimate interest to the mathematical statistician, opportunities for profitable employment of powerful computer facilities, and, above all, diverse requirements for collaboration and interaction between scientific experimenters and statisticians.

REFERENCES

1. Cochran, W.G. (1965). The Planning of Observational Studies of Human Populations. *Journal of the Royal Statistical Society, A128,* 234-266.

2. Colton, T. (1962). A Model for Selecting One of Two Medical Treatments. *Bulletin de l'Institute Internation de Statistique,* (*39*) (3), 185-200,

3. Grundy, P M., Healy, M.J.R., and Rees, D.H. (1956). Economic Choice of the Amount of Experimentation. *Journal of the Royal Statistical Society,* B18, 32-55.

4. Schwarz, D. and Lellouch, J. (1967). Explanatory and Pragmatic Attitudes to therapeutical Trials. *Journal of the Chronic Diseases 20,* 637-648.

IMPACT OF STERILISATION AND CONTRACEPTION ON FERTILITY

By

ALEYAMMA GEORGE, R. KRISHNA PILLAI AND Y.S. GOPAL*

This paper deals with the study of the effect of different hypothetical rates of sterilisation on the reduction in marital fertility**, assuming various levels of reduction in fertility due to contraceptive practices. The need for such a study arose when confronted with the question of suggesting a target rate of sterilisation of effectively married couples in order to bring down the fertility rate to a desired level within a specified period. For this study, the period is taken to be 1966-1976. The various rates of sterilisation to be adopted in order to reduce the fertility rate by desired amount, like 50 percent by 1976, can be obtained from the Tables. In order to have wider use, these reductions are provided for various assumed rates of fall in fertility due to contraception, including zero, that is, the case of no reduction due to *contraceptive practices.*

The data collected in various demographic studies conducted by the Department of Statistics, University of Kerala, have been utilised for computing the marital fertility rates for this study.

The various rates of sterilisation and reduction in fertility due to contraception used in this paper are more or less arbitrary but are thought to be reasonable and practicable in a population like that of Kerala. Approximate values for other rates, not used in this paper (but within the range considered here) could be obtained by interpolation. Of course it is admitted that the model taken is a simplified one when considering the complicated mechanism of fertility behaviour and the difficulties associated with the measurement of effective contraceptive practices ; however, it is hoped that this will serve as a guideline for those engaged in organising action programmes and for further research.

The Model

In order to study the fertility behaviour over a period of time, mainly two courses are open. One is to study the fertility history of cohorts of women over a number of years and on the basis of the results, make forecasts. The other is to make a cross section study at a particular instant of time and try to project the results to a period making certain assumptions which are reasonable. In the present study, the second approach has been

---

*Kerala University

**Throughout this paper fertility, unless otherwise stated refers to marital fertility.

used. Fertility reduction in this study refers to the reduction in age specific marital fertility rate. These rates for the different years 1967-76 have been calculated, making the following assumptions.

1. The proportion of unmarried women aged i getting married before age $(i+1)$ remains constant, from year to year. This is reasonable because there is no indication that there will be considerable change in the marriage pattern during the period under consideration, *viz.* 1967-76.

2. The probability of survival of women aged i to age $(i+1)$ remains constant from year to year. This assumption implies that there will not be much changes in mortality rates for women in the child bearing ages although this may not be fully justified.

3. The proportion of women becoming widows between age i and $(i+1)$ remains constant. Again the assumption is that the mortality pattern of men (husbands of the women in child bearing ages) continues to remain unchanged.

4. Under the present set-up, women who have three or more living children can opt for a sterilization operation. But for this study parity* has been taken as the criterion for opting for sterilization. It is assumed that the chance of a woman of parity 3 or more opting for sterilization is independent for her age or parity.

5. It is expected that there will be a constant annual addition to the number of people taking to contraception and it is, therefore, assumed that the additional reduction in marital fertility due to the contraceptive practices will remain constant, say δ. Thus if the present rate of reduction in fertility due to contraception is π, the reduction in marital fertility due to the increased contraception in the $(k+1)$th year will be $\pi+k\delta$. This effect is assumed to be 0·0025. Data were not available to get an estimated value of δ and considering the small additional proportion of people taking to contraception, this arbitrary value of 0·0025 is deemed satisfactory.

6. Since there are only few cases of parities higher than six, they were included in parity 6 in the study. Thus the parity 6 in the study actually refers to parity six or more. The following notations are used in the model :

$k=$ years after 1966, *i.e.,* years $1966+k$, $k=0, 1, 2, \ldots 10$.

$u_i^k=$ number of unmarried (single) women aged i in year k.

$m_{ij}^k=$ number of married women aged i and of parity j in the year k.

---

*Throughout this paper parity refers to the number of live births the woman had.

$f_{ij}{}^{k}$ = the age parity specific marital fertility rate for age i and parity j in year k.

r_i = proportion of unmarried women of age i getting married between age i and $i+1$.

P_i = probability of survival of a woman aged i to age $i+1$.

s_i = proportion of married women aged i becoming widows between age i and $i+1$.

π = reduction in the marital fertility rate of fertile women due to practice of contraception during the initial year, 1966 $(k=0)$.

$\delta = (0{\cdot}0025)$, the constant additional annual reduction in marital fertility rate due to contraceptive practices.

θ = *the proportion of married women in the child bearing* period and of parity 3 or more undergoing sterilization in a year.

Then the following relations are self evident :

$$f_{ij}{}^{k} = \begin{cases} f_{ij}{}^{k-1}\left[\,1-\overline{\pi+\delta(k-1)}\,\right], j<3,\ k\geqslant 1\ldots\ldots \\ f_{ij}{}^{k-1}\left[\,1-\overline{\pi+\delta(k-1)}\,\right](1-\theta), j\geqslant 3,\ k\geqslant 1 \end{cases} \qquad \ldots(2{\cdot}1)$$

The factor $(1-0)$ is to be introduced in addition for $j \geqslant 3$, as sterilisation also affects reduction. With the help of the above relations, age parity specific marital fertility rates could be worked out if $f_{ij}{}^{0}$ for the initial year are known. These initial age parity specific marital fertility rates were obtained from the data from the Family Planning Communication and Action Research Evaluation Study 1[3].

By considering the various possibilities during the immediately preceding year, the number of married women in the different age parity groups for the different years $k=1,\ldots 2,\ .\ 10$, could be obtained from the distribution of the initial year $k=0$ by means of the following expressions :

$$m_{ij}{}^{k} = \begin{cases} [m^{k-1}_{i-1,\,j}\ (1-f^{k-1}_{i-1,\,j} + u^{k-1}_{i-1}\, r_{i-1}](P_{i-1}-s_{i-1})^{*},\ \text{if } j=0 \\ [m^{k-1}_{i-1,\,j-1}\ f^{k}{}_{i-1,\,j-1} + m^{k-1}_{i-1,\,j}\ (1-f^{k-1}_{i-1,\,j}\ )](P_{i-1}-s_{i-1}), \\ \qquad\qquad 1\leqslant j<6\ ;\ k\geqslant 1 \\ [m^{k-1}_{i-1,\,j-1}\ f^{k-1}_{i-1,\,j-1} + m^{k-1}_{i-1,\,j}](P_{i-1}-s_{i-1}),\ j=6 \end{cases} \qquad \ldots(2{\cdot}2)$$

(the last case $j=6$ actually includes parities 6 or more). The above expression for $j=0$ involves $u_i{}^{k}$, the number of unmarried women of age i in year k

---

*As there were no data available for a woman getting married at a particular age and becoming widow before attaining the next age, and as the rate of widowhood itself is small, the same rate s_i was taken as an approximation for a woman becoming widow at the same age as age of her marriage.

for the various years. It could easily be seen that these can be calculated using the following recurrence relation.

$$u_i^k = u_{i-1}^{k-1}\ P_{i-1}(1-r_{i-1}),\ k \geqslant 1 \qquad \ldots(2{\cdot}3)$$

making use of the initial values u_i^0 for various ages i.

After finding the f_{ij}^k's and the m_{ij}^k's as defined by (2·1) and (2·2), the age specific marital fertility rates were obtained as the ratio of the total number of births during that year to the married women of a particular age group to the number of married women in that age group. Then the age specific marital fertility rate of age i during year k is given by

$$f_i^k = \frac{\sum\limits_{j=0}^{6} m_{ij}^k f_{ij}^k}{\sum\limits_{j=0}^{6} m_{ij}^k} \times 1000 \qquad \ldots(2{\cdot}4)$$

These age specific marital fertility rates are calculated on the basis of different combinations of the values of the rate of sterilization, $\theta = 0{\cdot}00125,$ ......, 0·03 (2·5) and different rates of fall in fertility due to contraception,

$$\pi = 0{\cdot}0125, \ldots\ldots, 0{\cdot}3 \qquad \ldots(2{\cdot}6)$$

The age distribution of women according to marital status was obtained from the 1961—Census data for Kerala [1]. As the census data give only quinquennial age grouping for the distribution, the data has been redistributed in order to arrive at single year ages, by using Sprague multiplier. Further redistribution according to parity within each age has been accomplished by utilising data from the evaluation study I, [3]. The survey which was carried out in 1965 contained the age parity information for women. Age heaping in the data of that study has been eliminated by smoothing and this final distribution provided the basis for redistribution of women in the single year ages according to parity.

The age parity specific marital fertility rates for 1966 given in Table 1 have also been calculated from the data of the evaluation study I, [3]. Actually these fertility rates refer to the period 1963-64, but in the absence of other reliable figures or methods of adjustment, these rates themselves are taken as the initial rates, for 1966, in the present study.

The marriage rate r_i and the widowhood rate s_i have been calculated using the data from the survey conducted in the coastal area of Quilon district during 1965, [2]. The data included the age distribution according to marital status. The mariage rate r_i of a woman aged i getting married before age $i+1$ has been taken as the ratio of the number of married women whose age at marriage was i, to the total number of women aged i. Similarly, the widowhood rate s_i, is taken as the proportion of married women aged i getting widowed before age $i+1$; it is the ratio of the total number of married women who have become widows at age i to the total number of married women of age *i*. It was felt that these rates based on the more recent Quilon Study [2] will be more up-to-date than those obtainable from the census of 1961, [1]. Moreover, the Census of 1961 data did not provide information on age at marriage nor age at widowhood.

The survival rates for the present study has been obtained by making use of the model life tables published by the United Nations, [7]. Rates for females were taken from the table corresponding to a life expectancy of 50 years. The 5 year rates given in the model life table were then interpolated to obtain single year survival rates.

The various marital fertility rates obtained for the different years from 1967 to 1976 are presented in Tables 1 to 14.

Table 1 gives the summary findings, namely the total marital fertility rate for 1976 for the values θ and π combinations. The initial total marital fertility rates (that is for the year 1966) was 270·3. As one would expect, the reduction in the rate becomes more and more when either the reduction due to contraception is increased or when sterilisation rate is increased. However, this fall is much more sharp when reduction due to contraception is enhanced than when rate of sterilisation is increased. Table 1 can also be used to set up a target sterilization rate in order to achieve a target reduction in the marital fertility rate by 1976, in a community of the type under study with known rate of reduction due to contraceptive practice.

TABLE 1

Total marital fertility rates for different combinations of π and θ in 1976.
(Figures in brackets give crude birth rates)

θ π	*0·00125*	*0·0025*	*0·005*	*0·01*	*0 02*	*0·03*
0	256·9 (35·2)	255·4 (35·1)	252·5 (34·7)	236·7 (34·0)	246·0 (32·8)	226·1 (32·2)
0·0125	205·5 (29·0)	204·3 (28·8)	201·7 (27·5)	197·8 (23·0)	189·6 (27·1)	186·0 (26·2)
0·025	180·4 (25·9)	180·0 (25·8)	178·2 (25·6)	175·7 (25·3)	168·7 (24·4)	162·2 (23·6)
0·05	142·1 (21·0)	141·4 (20·9)	140·0 (20·7)	138·0 (20·5)	132·5 (19·8)	128·2 (19·2)
0·1	83·7	83·4	82·6	81·5	78·4	76·8
0·2	25·8	25·7	25·4	25·2	24·7	24·1
0·3	6·7	6·7	6·6	6·5	6·4	6·[illegible]

Initial rate 270·3 (40)

Table 2 gives the total marital fertility rates for the ten year period 1967-76, according to the various rates of reduction due to contraception and sterilization. It can be seen that when a 5 per cent reduction due to contraception can be achieved, the rate of sterilization, in the range considered is not much important, as there is an approximate 50 per cent reduction in the marital fertility rate and crude birth rate for all rates of sterilisation considered. However, this rate of reduction due to contraception may perhaps be too ambitious one for the present day conditions in India. If 30 sterilizations can be done for every 1000 married women who are currently married, we can expect at least 16 per cent reduction in the marital fertility rate by 1976.

TABLE 2

Total marital fertility rate according to reduction in birth rate due to contraception (π) 1967-1976.

	Year π	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
θ=·00125	0·0125	260·6	252·1	245·2	239·4	234·1	229·1	223·3	217·5	212·0	205·5
	0·025	257·3	246·2	236·3	227·9	220·3	212·7	205·3	193·6	190·2	180·4
	0·05	250·7	233·9	219·0	206·1	194·4	183·2	172·4	162·3	152·1	142·1
0=·0025	0·0125	260·1	252·0	244·8	238·8	233·3	228·0	222·5	215·9	210·8	204·3
	0·025	256·9	245·8	235·9	227·3	219·6	212·0	204·5	191·1	189·2	180·0
	0·05	250·6	233·6	218·6	205·6	193·8	182·6	171·8	156·8	151·4	141·4
0=·005	0·0125	259·3	251·4	243·9	237·2	231·7	226·1	220·5	213·4	208·6	201·7
	0·025	256·5	244·9	234·9	225·9	218·2	208·6	202·8	188·0	185·2	178·2
	0·05	250·4	233·0	217·7	203·1	192·6	181·3	170·5	155·4	150·0	140·0
0=·01	0·0125	259·5	250·1	242·0	235·1	229·0	223·0	217·0	204·7	204·6	197·8
	0·025	256·2	244·0	233·2	223·9	215·5	207·4	199·5	185·7	183·8	175·7
	0·05	250·3	233·0	216·2	202·5	190·3	178·7	167·9	152·5	147·3	138·0
θ=·02	0·0125	258·1	247·6	288·3	230·4	223·2	216·4	209·9	197·4	194·6	189·6
	0·025	254·9	241·1	229·8	219·5	210·4	201·7	193·3	179·7	177·0	168·7
	0·05	248·4	229·4	212·9	198·5	185·7	173·9	162·8	149·8	142·2	132·5
0=·03	0·0125	256·4	245·1	234·7	225·9	218·0	210·5	203·4	190 5	189·6	185·8
	0·025	254·0	239·1	226·2	215·1	205·2	196·0	187·3	173·7	170·0	162·2
	0.05	247·1	227·1	209·7	195·1	181·8	169·2	158·3	143·4	137·8	128·2

Tables 3-14 present the age specific marital fertility rates for the ten year period 1967-1976 for the various combinations of the different rates of sterilisation and the different rates of reduction due to contraception. In Table 3-5, the sterilisation rate varies but the reduction due to contraception remains constant, namely 0·0125.

TABLE 3

Age specific marital fertility rate, 1967-76.

π=0·0125, sterilization rate θ=·00125

Year Age	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	394·6	386·8	379·2	371·5	364·6	356·2	347·2	338·2	327·3	316·7
20—24	361·5	350·6	345·5	342·7	338·9	332·6	325·7	318·1	310·0	300·5
25—29	298 6	294·0	287·7	281·7	275·9	270·9	261·9	255·9	248·1	238·3
30—34	230·8	226·5	221·1	216·0	209·8	205·3	190·0	194·4	188·6	181·2
35—39	165·7	162·1	158·7	155·0	151·2	145·6	137·6	130·3	126·3	123·4
40—44	95·6	95·1	92·9	91·5	89·2	86·6	84·1	81·4	77·9	73·3
All ages	260·6	225·1	245·2	239·4	234·1	229·1	223·8	217·5	212·0	205·5

π=Reduction in birth reduction due to contraception in 1966.
θ=Rate of sterilization.

TABLE 4

π=**0·0125 Sterilization rate** θ=**0·01**

Year Age	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	395·4	386·8	379·2	371·6	364·7	356 2	347·2	338·4	327·3	316·7
20—24	360·9	349·5	344·0	341·0	337·0	330·4	323·4	315·7	307·5	297·9
25—29	297·0	291·0	283·1	275·6	268·3	260·4	252·1	244·0	236·3	226·9
30—34	229·1	223 1	210·3	209·6	204 0	196·2	190·0	182·7	217·6	167·0
35—39	164·3	159·4	154·7	150·0	144·8	138·3	131 8	125·0	119·0	113·8
40—44	94·8	93·4	90·4	88·4	85·4	78·6	77·9	75·7	71·7	67·1
All ages	259·5	250·1	242 0	235·1	229·0	223·0	217·0	210·9	203·6	197 8

TABLE 5

π=**0·0125 Sterilization ratio** θ=**0·03**

Year Age	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	395·4	386·8	379·2	371·6	364 7	356·2	347·2	338·4	327·3	316·7
20—24	359·6	347·2	340·9	337·3	332·8	325·8	318·5	310 6	301·9	292·5
25—29	293·4	283 9	272·9	262·2	252·0	241·5	231·3	221 9	213·4	203·8
30—34	225·1	215·6	205·4	195·7	186·2	176·8	168 2	159 0	149 6	140·1
35—39	161·1	153 3	146 0	138·5	131·2	122·9	114·7	106 6	99·3	93·0
40—44	92 9	89·6	85·0	91·4	77·4	68 4	68·2	64·2	59·2	54·0
All ages	256·5	245·1	234·7	225·9	218·0	210·5	203·4	196·3	189 6	185·0

The above tables reveal that the age specific marital fertility rate shows a steady decline as the rate of sterilization increases. The same pattern is indicated by the gross marital fertility rate. The maximum reduction in age specific marital fertility rate occurs in the age group 35-39, showing a decline of 27·3 per cent by 1976 when the sterilization rate 0 = 0·00125, going down to 45·2 per cent by 1976 the sterilization rate is 0·03. This maximum reduction occurring in this age group 35-39 may be an indication that more sterilizations are likely to take place in this age group as people may be feeling it too early to do it before and not worth doing when the wife is on the wrong side of 40. Such a conclusion gains more strength in the light of the fact that the reduction in the age specific marital fertility for the group of women in ages 40-44 is consistently higher than that in the age group 35-39.

The next three tables, tables 6 to 8, give the age specific marital fertility and gross marital fertility for 1967-76, for various rates of sterilization, with reduction due to contraception 0·025.

TABLE 6

$\pi = 0·025$ Sterilization rate $\theta = 0·00125$

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	390·4	377·2	365·4	353·9	343·2	331·3	319·1	307·3	293·6	280·5
20—24	256·9	342·1	333·4	327·4	320·4	311·3	301·6	291·9	281·2	270·1
25—29	294·8	286·7	277·1	268·0	259·3	250·0	240·5	231·1	221·8	211·1
30—34	227·9	220·8	212·8	205·0	196·8	190·7	183·0	179·5	167·1	158·2
35—39	170·5	158·0	152·8	147·3	141·7	134·7	127·7	124·9	114·1	108·6
40—44	94·4	92·6	89·4	87·0	83 6	80·0	76·6	73·2	69·0	64·2
All ages	257·3	246·2	236·3	227·9	220·3	212·7	205·3	198·9	190·2	180·4

TABLE 7

$\pi = 0·025$ Sterilization rate $\theta = 0·01$

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	390·4	377·2	365·4	353·9	343·2	331·3	319·1	307·3	293·6	280·5
20—24	356·3	341·1	332·1	325·8	318·7	209·5	299 9	289·9	679·3	268·1
25—29	293·3	283·7	272·7	262·3	252·3	242·0	231·7	221·7	212·4	201·7
30—34	226·2	217·6	208·1	199·2	189·5	181·7	173·6	165·0	156·1	146·8
35—39	262·2	155·4	149·0	142·4	135·8	128·0	120·2	112·5	105·5	99·5
40—44	93·6	91·0	87·0	83·9	79·0	75·9	71·9	68·1	63 6	58·5
All ages	256·2	244·0	233·2	223·9	215·5	207·4	199·5	191·4	183·8	175 7

TABLE 8

π=0·026 Sterilization rate θ=0·03

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	390·4	377·2	365·4	353·9	343·2	331·3	319·1	306·5	293·6	280·5
20—24	355·0	388·8	329·1	322·3	314·9	305·5	295·8	285·7	275·1	264·0
25—29	289·7	276·9	262·9	249·6	237·1	224·8	213·0	202·4	192·9	182·6
30—34	229·1	210·2	197·7	186·0	173·8	163·9	153·9	143·8	133·6	123·5
35—39	159·8	149·5	140·5	131·6	123·1	131·7	104·5	96·0	88·2	81·4
40—44	91·7	87·3	81·8	77·7	72·2	67·0	62·2	57·6	52·8	47·2
All ages	254·0	239·1	226·2	215·1	215·1	196·0	187·3	179·0	170·9	162·0

These Tables indicate the same pattern as exhibited by Tables 3-5 for the rate of reduction due to contraception 0·0125. The maximum reduction in age specific marital fertility occurs in the age group 35-39, reduction rising from 36·0 per cent for 0·00125 sterilization to 52·1 per cent when sterilization rate is 0·03. There is a steady fall in the annual gross marital fertility, bringing about a 40 per cent reduction by 1976.

Tables 9 to 11 provide the age specific marital fertility and gross marital fertility for *1967-1976* for the various sterilisation rates considered, assuming the rate of reduction in fertility due to contraceptive practices to be 0·05.

TABLE 9

π=0·05 Sterilization rate θ=0·00125

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	380·4	358·5	338·9	320·3	303·1	285·8	268·3	251 2	234·5	218·2
20—24	347·7	325·4	310·1	298·0	285·5	271·7	257·9	243·9	229·9	215·8
25—29	287·3	272·3	256·6	242·1	228·5	214·9	201·7	189·0	177·2	164·6
30—34	222·0	209·6	197·4	185·3	172·8	163·6	152 3	142 0	131·7	121·3
35—39	159·4	150·0	141·3	132·7	124·2	114·9	105·9	97·1	89·3	82·5
40—44	92·0	87·9	82·6	78·3	73·3	68·2	64·3	59·0	54·0	48·7
All ages	250·7	233·9	219·0	206·1	194·4	183·2	172·4	162·3	152·1	142·1

TABLE 10

π=0 05 **Sterilization rate** 0=0·01

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	380·4	358·5	338·9	302·3	303·1	285·6	268·3	251·2	234·5	218·3
20—24	347·2	324·4	308·9	296·6	284·1	270·3	256·5	242·6	228·6	214·7
25—29	285·7	269·4	252·6	237·0	222·5	208·3	194·8	182·1	170·4	158·1
30—34	222·0	206·5	192·5	179·6	166·5	155·5	144·7	133·9	133·5	116·0
35—39	158·1	147·5	137·8	128·2	119·1	109·2	99·7	90·7	82 6	76·6
40—44	91·1	86·3	80·4	75·6	70·1	64·7	59·6	54·8	49·7	44·5
All ages	250·3	233·0	216·2	202·5	190·3	178·7	167·9	157·2	147·3	138·0

TABLE 11

π=0·05 **Sterilization rate** 0=0·03

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	380·4	358·5	338·9	320·8	303·1	285·6	268·3	251·2	234·5	218·3
20—24	345·9	322·3	306·1	293·6	281·1	267·2	253·5	239 9	226·1	212·4
25—29	282·2	263·0	243·6	227·9	211·5	194·1	180·0	167·5	156·5	146 2
30—34	216·6	199·6	182·9	167·8	152·1	142·1	130·0	116·8	107·2	96·6
35—39	155·1	141·9	129·5	118·6	107·9	97·1	86·6	77·7	69·2	52·1
40—44	89·6	82·9	75·6	69·6	63·2	57·1	51·5	46·3	41·0	35·8
All ages	247·1	227·1	209·7	195·1	181·8	169·2	158·3	147·9	137·8	128·2

The findings of these Tables are also consistent with the previous two sets, Tables 3 to 5 and Tables 6 to 8. The notable feature is that the reduction in the age specific marital fertility rates are considerably high, compared with the previous two sets. Here, again, the reduction is maximum more or less among the age group 35-39, but however, there is not much appreciable change according to the rate of sterilization. The gross marital fertility shows a reduction of 47·4 per cent by 1976 with rate of sterilization 0 00125 and this reduction in gross marital fertility becomes 52·6 per cent by 1976 when sterilizatian rate is 0·03.

When the reduction due to contraception is only 0·0125 the total marital fertility rate for 1976 for the different sterilization rates ranged from 185 to 205·8, a reduction of 24 to 31·6 per cent from that of 1966, but it ranges from 162·2 to 180·4, a reduction of 33·3 per cent to 40 per cent when the reduction due to contraception is assumed to be 0·025. Further, the rate of reduction rapidly increases with increased reduction due to contraception. This seems to suggest that the more important factor can be reduction due to contraception than sterilization within the range considered in this paper. Perhaps, with higher rates of sterilisation than those considered here, this dependence may change direction but it cannot be forgotten that even the highest rate of sterilization considered in this paper, namely 30 per thousand currently married couples is too much to achieve in practice. It can be seen that the total marital rates very rapidly fall for the higher rates of reduction due to contraception considered. Of course, one should not be too much attracted by this because to achieve that rate of reduction due to contraception may be even more difficult than achieving a little more increase in the rate of sterilization.

The above fact is more clearly brought out by Tables 12 to 14 which give the age specific marital fertility rates for the period 1967-76 for various rates of sterilization assuming no reduction due to contraception ($\pi = 0$, $\delta = 0$).

TABLE 12

Sterilization rate $\theta = 0·00125$

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	400·4	397·4	396·1	395·6	396·4	396·4	396·4	396·4	396·4	396·4
20—24	266·0	360·0	360·0	362·7	366·1	367·0	367·6	367 8	367·5	366·3
25–29	302·4	302·2	300·8	299·9	300·5	300·3	299·8	299·7	300·2	300·2
30—34	233·7	232·5	231·3	229·8	229·0	229·8	230·4	230·4	230·0	229·1
35—39	167·8	141·3	166·2	165·6	165·2	163·4	161·7	159·9	159·0	159·5
40—44	96·8	97·7	97·2	97·8	97·6	96·9	97·1	97·7	96·5	94·9
All ages	265·7	259·4	256·0	255·0	254·8	254·9	255·4	256·1	256·5	256·9

TABLE 13

Sterilization rate $\theta = 0·01$

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15—19	400·4	397·4	396·1	395·6	396·4	396·4	396·4	396·4	396·4	396·4
20—24	365·4	359·0	358·1	361·8	364·2	363·6	364·6	364·5	363·8	362·6
25—29	300·8	289·0	295·9	293·8	292·2	290·2	288·2	286·6	285·7	284·8
30—34	232·0	229 4	225·4	223·0	220·3	219·6	218·4	216·7	214·9	212·0
35—39	166·3	163·9	162·0	160·1	158·3	155·2	152·2	149·1	146·9	146·0
40—44	96·0	96·0	94·7	94·4	93·3	92·1	91·2	90·4	88·7	86·4
All ages	264·5	256·7	252·9	250·5	249·2	248-3	247·8	247·6	246·7	246·7

TABLE 14

Sterilization rats $\theta = \cdot 03$

Year *Age*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*	*1975*	*1976*
15-19	400·4	397·4	396·1	395·6	396·4	394·4	396·4	396·4	396·4	396·4
20-24	364·1	356·6	355·5	357·2	359·4	358 9	358·4	357·6	356·1	354·1
25-29	297·1	291·8	285·2	279·4	274·1	268·6	263·5	259·2	255·9	252·7
30-34	228·0	221·7	214·9	208·2	202·1	197·8	193·1	188·1	182·8	177·4
35-39	163·2	157·6	152·7	147·9	143·4	137·8	132·3	127·0	122 4	118·9
40-44	94·0	92·1	89·2	86·9	84·1	81·3	78·7	76·3	73 2	69·6
All ages	261·6	251·9	245·3	240·5	236·9	233·5	231·9	229·1	228·6	226·1

The values given in the last row of each table indicate the pattern of fall in total marital fertility. Even with the highest rate considered namely 30 sterilization per 1000 effectively married women, the total marital fertility rate of 273 for 1966 falls only to 226 by 1976, a fall of only sixteen per cent. These total marital fertility rates indicate a considerable dependence on the sterilization rate assumed.

5. Summary

An attempt has been made here to evaluate possible changes in the marital fertility for the ten year period 1967-76 which could be brought about by adopting various rates of sterilization for married women who already have three or more live births assuming various rates of reduction in fertility due to contraception. It is seen that dramatic changes in the marital fertility rate can be brought about by the adoption of different rates of sterilisation coupled with different reduction due to contraception. It is also seen that while the reduction in the marital fertility rate is highly sensitive to changes in the rate of sterilisation, it is more so for the changes in the reduction due to contraceptive practices. It is hoped to supplement these studies with modified rates of fertility obtained on the basis of continuous periodical surveys. Such studies will throw some light on the mechanics of fertility reduction which could be achieved by adopting various rates of sterilization, a long felt filling up of a gap which exists in the fertility studies of a country like India.

In order to have a better understanding of the trend in fertility under the assumptions used in this paper, the calculations are being extended to cover a wider period of time than ten years. The results of these calculations will be soon presented as a supplement to this paper. ·

ACKNOWLEDGMENT

The main computations for the various tables presented in this paper were carried out by the computing section of the Thumba Equatorial Rocket Launching Station, Trivandrum. But for the Minsk Computer and the co-operation of the staff of the computing section there, this paper could not have been completed.

REFERENCES

1. Census of India 1961-Kerala.
2. Department of Statistics, University of Kerala: Demographic Study of the Monozite Area in Quilong District (1966).
3. Department of Statistics, University of Kerala: Evaluation Study I; Impact of Family Planning Education on Wives (1966).
4. Jaffe, A.T.: Handbook of Statistical Methods for Demography, U.S. Bureau of the Census, pp. 94-101 (1960).
5. Murthy, D.V.R.: "*Estimated reduction in birth rate resulting from different* combinations of sterilization and contraception programmes in India." Proceedings of world population conference, Belgrade (1965),
6. Mendel C. Sheps and Edward B. Perin: "Changes in Birth Rates as a function of contraceptive effectiveness: some applications of a Stochastic Model"; Amer. Jour. Public Health.—Vol. 53, No. 7, pp. 1031-1046 (1963).
7. United *Nations Department* of Social Affairs, Population Branch: Age and Sex Patterns of Mortality—Model Life Tables for Under Developed Countries: New York (1955).

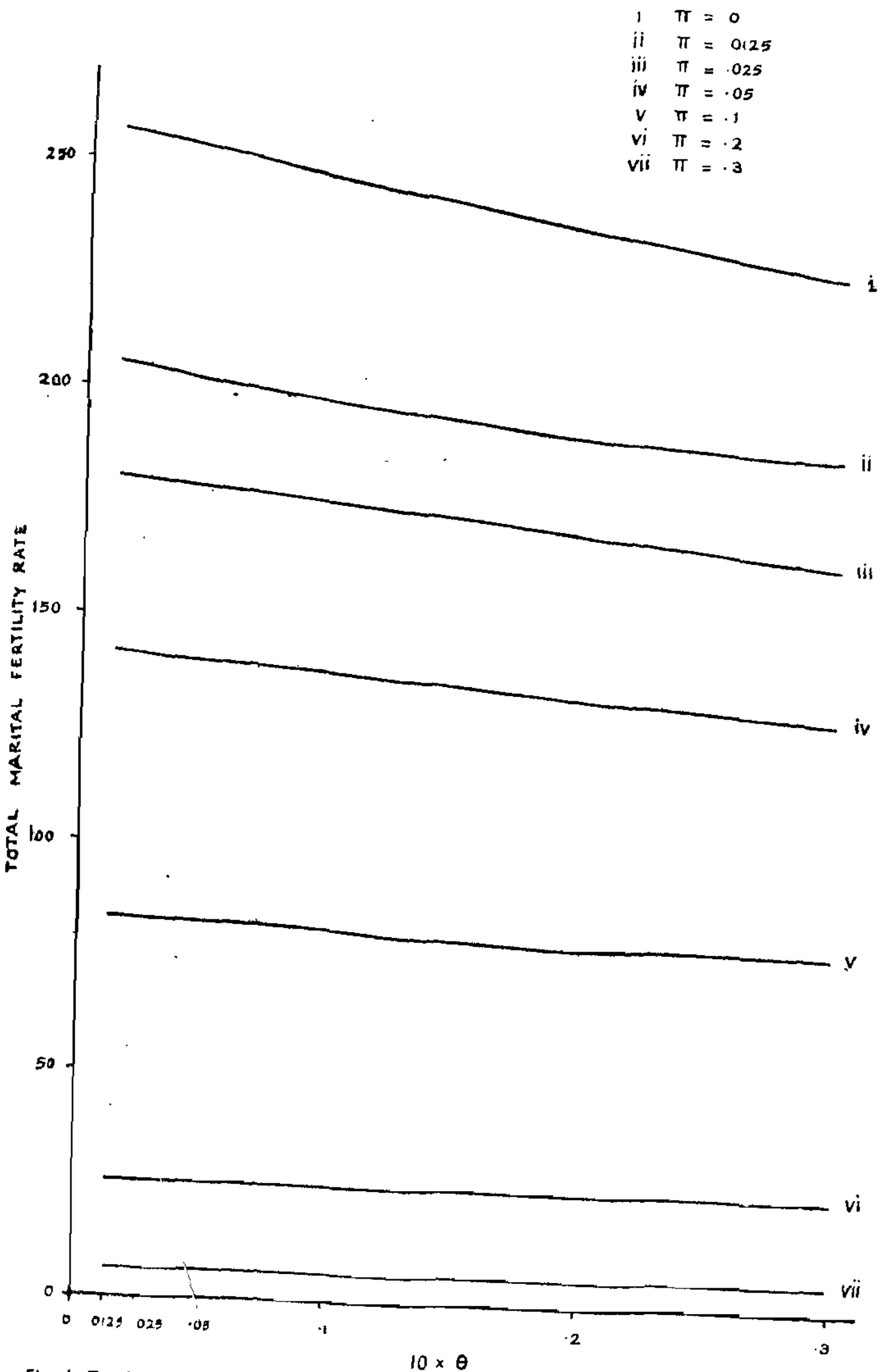


Fig. 1. Total Marital Fertility Rate according to Rate of Sterilization θ for various rates of reduction in fertility due to contraception π.

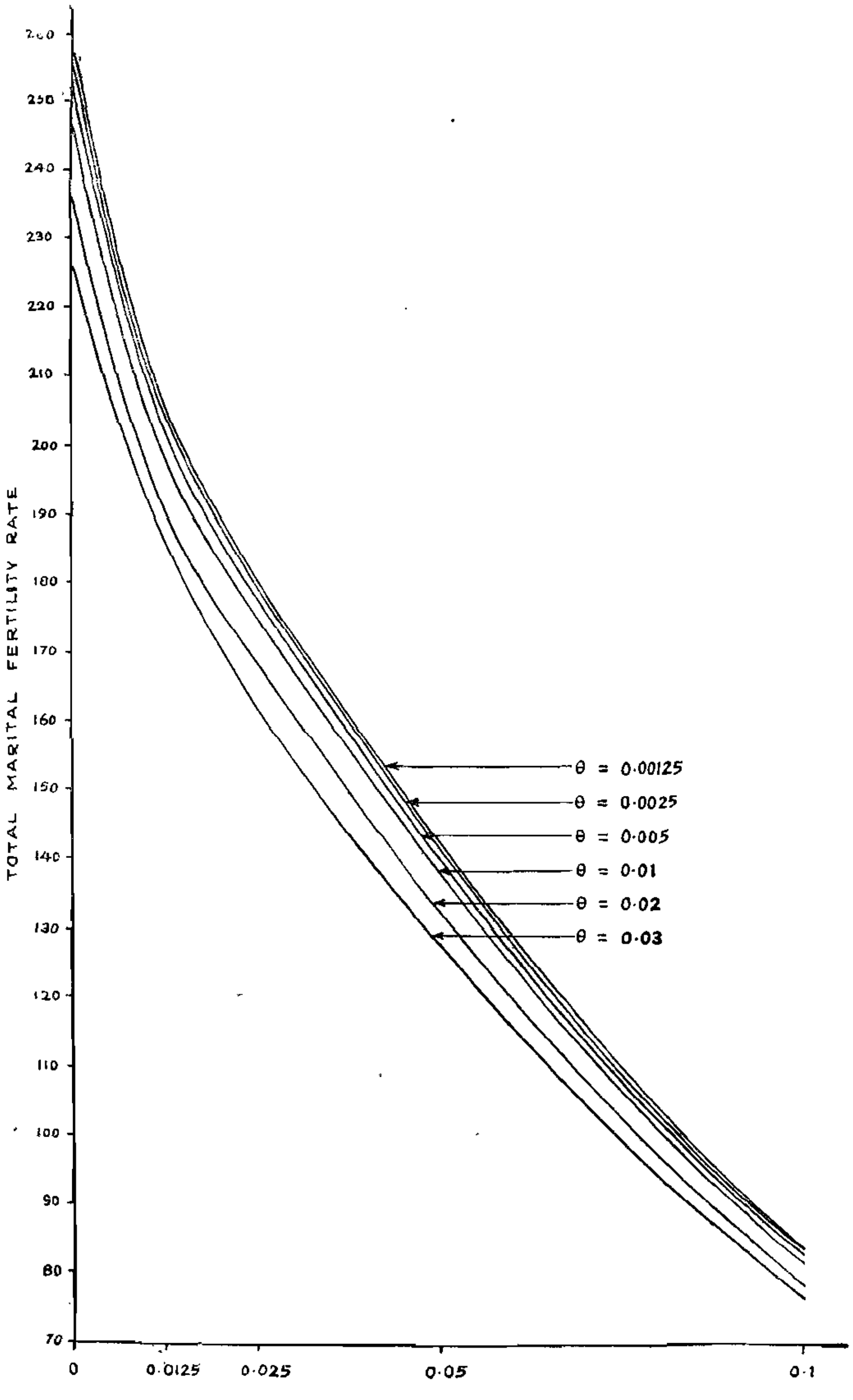


Fig. 2. Total Marital Fertility Rate according to Rate of Reduction in Fertility due to Contraception π, for various rates of Sterilization, θ

SOME RECENT IMPROVEMENTS IN AGRICULTURAL STATISTICS AND TASKS AHEAD

By

R. GIRI

During the last two decades, considerable headway has been made in securing increasingly reliable, comparable and comprehensive data to meet the needs of food and agricultural administration and policy decisions and formulation and execution of the agricultural development plans and their periodic assessment and evaluation. With the growth and diversification of agricultural economy and the requirements of future planning, further improvements are needed in the existing data. Some of these improvements are discussed briefly in this paper.

Extension of Reporting Area

As a result of extension of cadastral survey and institution of reporting agencies in certain parts of the country, "reporting area" based on complete enumeration has now increased to 77·0 per cent of its total geographical area. Estimates in respect of another 5 per cent have become available by sample survey. Estimates based on conventional methods are also framed for 12 per cent of the geographical area which lie mostly in Orissa, Rajasthan, Gujarat, Assam and Tripura. The extent of "non-reporting area" has now shrunk to 6 per cent, the bulk of which lies in Jammu and Kashmir, NEFA, Manipur and Himachal Pradesh.

The non-reporting area is covered mostly with dense forests and hills and is sparsely cultivated, where cadastral survey and enumeration of area will be both difficult and prohibitive from cost point of view. For such areas suitable sampling techniques like the one recently tried in the hill districts of U.P. may be adopted to frame the estimates. However, the sample survey in extensively cultivated areas of West Bengal, Orissa and Kerala which are also cadastrally surveyed and have now revenue and other agencies at the village level, do not meet the requirements of comprehensive data for planning and administrative purposes. Considering the importance of dependable statistics at district and block levels for both planning and operational purposes, it would be desirable to examine the feasibility of utilising village level agencies set up recently in these States, for building up la statistics on complete enumeration basis.

In the States in which area statistics are already based enumeration, the jurisdictions of primary reporting agencies

excessive in view of multifarious duties they are called upon to perform in the context of welfare and development activities of the State and democratisation of administration. Their charges and correspondingly the charges of supervisory officers need to be appropriately reduced by suitably augmenting their strength.

The land records manuals in different States make adequate provision for supervision by the superior officers over the work of area enumeration by primary reporters. But the villages and fields therein are generally selected for supervision by rotation according to a certain roster which gives an idea to the primary reporters of the villages unlikely to be visited in a particular year. An element of surprise can be introduced in the whole supervision programme by randomly selecting villages and fields for supervision. This would also serve another important purpose, in the sense that it would provide an assessment of the extent of unreliability of area enumeration due to neglect of work by the primary reporter under pressure of work. A part of this supervision can be entrusted to independent statistical agencies.

Adoption of Standardised Concepts, Definitions and Classifications

Consequent on the divergent systems of maintenance of land records, the emerging area statistics are not strictly comparable from State to State unless certain basic uniformity is enforced. To meet this need, the Technical Committee on Coordination of Agricultural Statistics (TCCAS) set up in 1949 by the Union Ministry of Food and Agriculture recommended standard basic and abstract village forms, uniform concepts and definitions and standardized classifications.

The Standing Committee on Improvement of Agricultural Statistics (CIAS), set up in the Ministry of Food, Agriculture, Community Development and Cooperation, has, in recent years, examined and suggested revisions in the forms of almost all the States without disturbing their basic structures, with a view to ensuring essential conceptual uniformity and complete coverage of the necessary items. The land record forms generally follow the land classification adopted at the time of settlement ; and to enable culling out of comparable land-use data, the CIAS has worked out for each State necessary equations between the standard categories and the local land descriptions.

The distinctions between mixed-sown area and multiple-sown area is often not clearly understood, and the procedures followed in the different States for recording of area under mixed crops and allocation of net areas to component crops widely differ and lack objectivity. The proper accounting of areas under vegetables and other crops sown in one season but harvested in successive seasons, presents special problems. Improvements are also needed with regard to the recording of area under fruit orchards which are planted both pure and mixed and often have ground cultivation. There are several other problems like recording and proper accounting of areas under various types of bunds, recording of areas of crops like cardamom and pepper cultivated in forests and fruit orchards, etc. which need consideration from the point of view of improving the reliability and comparability over space and time, of area statistics. The CIAS is engaged in tackling these problems and has made several recommendations which the States are implementing.

Crop-cutting Surveys

The crop-cutting surveys by the random sampling method have been gradually extended to more crops and areas during the last two decades. In 1964-65, about 96 per cent of the production of cereals, 70 per cent of pulses, 94 per cent of sugarcane, 74 per cent of oilseeds, 80 per cent of cotton and 99 per cent of jute were covered by the crop-cutting surveys. The crops with regard to which the coverage now needs to be improved are small millets like ragi and kodakutki ; pulses like urad, moong, masoor and lytharus ; and oilseeds like sesamum, rape and mustard, linseed and castorseed.

The position with regard to production estimates of fruits and vegetables and minor crops of commercial importance is also not satisfactory. Only ad-hoc estimates are available for some of these crops after a time-lag of two to three years. The Institute of Agricultural Research Statistics (IARS) has conducted pilot investigations in a few typical districts of selected States with a view to evolving suitable sampling techniques for estimation of area and output of different fruit crops and for collection of data on their cultivation practices. In order to obtain reliable estimates of production of fruits, vegetables and commercial crops, it is necessary to extend the yield estimation surveys on those of the crops for which suitable sampling techniques have been evolved, to all the principal areas growing them. In the case of other crops on which sample surveys are either not initiated or have not yielded conclusive results, research work to finalise the sampling technique needs to be intensified.

The quality of the data collected from the crop-cutting surveys is largely dependent on the extent of supervision exercised over the field work. The departmental and statistical officers of the State departments concerned with these surveys and the NSS supervisory staff have to carry out extensive inspections of the field work. The response with regard to conduct and supervision of the experiments is, however, not uniform in all the States. There is considerable room for improvement in many States which is possible with the strengthening of their primary and supervisory agencies and the statistical staff.

Crop Forecasts

With improvements in area and yied statistics, the scope of crop forecasts has been gradually extended to more crops and larger areas. In 1947-48 when the Directorate of Economics and Statistics in the Ministry of Food and Agriculture took over Crop Forecasts from the Directorate General of Commercial Intelligence and Statistics, forecasts were issued for only 13 crops. To these crops have been added till to-day 21 more crops. Proposals for issue of forecasts for 9 more crops have been formulated. In addition to these crops, ad-hoc estimates are also prepared for some crops.

Completion of crop inspection in all the villages under the charge of a primary reporter and compilation of data at successive levels thereafter, take considerable time, and the area statistics based thereon become available long after the close of the agricultural year, with the result that they cannot be used for current planning and administrative purposes and for the purpose of crop forecasts. The area figures used for these purposes are naturally subjective estimates. More raliable figures can be obtained in time if the primary reporter is required to complete the crop inspection and compilation of the data collected, in respect of different villages in different months according to a phased programme. Some work in this connection has been done in the States of U.P. and Madhya Pradesh and this may provide guide-lines for other States to follow.

Certain methodological studies are also called for to improve the value of forecasts of long-duration crops like cotton, sugarcane, etc. the yield of which is obtained in a number of harvests. Suitable methods to estimate the total yield on the basis of some initial harvests only, need to be evolved to improve the reliability and timeliness of the forecasts of such crops. The usefulness of the forecasts can be enhanced if the yield estimates of the different varieties of a crop harvested at different times of the year are prepared, soon after their harvest without waiting for the harvest of all other varieties.

Irrigation Statistics

The reliability and inter-State comparability of statistics of crop-wise and source-wise irrigated areas as contained in Land Utilisation Statistics, are affected for want of adequate and clear provision for recording of the necessary irrigation particulars in the land record forms. Lack of adoption of uniform concepts regarding irrigated areas and of standardized classification of irrigation sources also affects the accuracy and inter-State comparability of irrigation statistics. The village-level abstracts of crop-wise and source-wise irrigated areas and of a number of different classes and sub-classes of irrigation sources are not prepared in some States. The CIAS has, therefore, suggested some revisions in the land record forms and Manuals of the States to improved the reliability and inter-State comparability of irrigation statistics iin India. Some of these revisions have been introduced in the course of the last six years.

Besides Land Utilisation Statistics, another source of statistics in regard to development of irrigation resources, is the progress reports furnished periodically by State Governments on irrigation schemes. The progress reports on major-medium irrigation projects give annual progressive totals of irrigation potentials created and utilised. The progress reports on minor irrigation schemes furnish the gross area benefited.

The areas benefited by minor irrigation reported in the progress reports do not represent correctly the additions to irrigated areas, for they include: (*a*) old irrigated area over which irrigation has been made more certain as a result of renovation of old works; (*b*) area benifited by water conservation-cum-ground water recharging schemes; and (*c*) area benefited by drainage, flood control, etc. The figures reported in the progress reports relate to gross irrigation potential in some cases and to net irrigation potential in others. Further in some cases, the additional areas benefited might be indicated, while in others cumulative totals might be given.

In order to improve the quality of irrigation statistics furnished in progress reports, it is necessary to appropriately classify the areas benefited from the different types of schemes. As the accuracy of the figures of benefits reported for the individually owned private works depends on the correctness of the yardsticks adopted, these yardsticks need to evolve on a realistic basis, with due regard to regional variations.

Sample surveys have been conducted by the Directorate of National Sample Survey (NSS) to ascertain the number of minor irrigation works functioning out of those reported to have been constructed, the extent and pattern of expenditure incurred, and the area irrigated. These surveys, however, have not covered all the States simultaneously and continuously over a period, thereby precluding a year-to-year appraisal of irrigation programmes. They have, however, highlighted the need for organising continuous comprehensive surveys of irrigation projects for assessing the additional area irrigated and for studying the other related aspects like change in cropping pattern, increase in cropping intensity, and improvement in yield-rates of crops.

Benefits of Improved Agricultural Practices

Estimates of areas brought under various improved practices is necessary to assess the impact of development measures in agriculture. From 1958-59, the NSS Directorate has conducted sample surveys covering improved seed, chemical fertilizers and chemical pesticides. The reaction of cultivators to various aspects of improved practices is also ascertained. Most of the data are collected by the enquiry method, although spot-check is carried out in respect of about 25 per cent of the fields in the selected holdings. It has, however, been difficult to verify the application of fertilizers and manures by direct observation. Proper identification and assessment of the extent of purity of improved seeds has also been difficult to assess.

The States generally furnish lumped figures for areas under improved seeds of various crops and the reported figures are not reliable. The yardstick concept, as at present adopted for assessing the additional production through the use of improved seed, also does not give reliable estimates. It is, therefore, necessary that the NSS surveys should be extended to cover the entire area in each State and all important improved practices including improved seeds. With regard to soil conservation, data are needed for studying production as a function of soil-kind, soil conservation system and time-spread after adoption of a soil conservation measure and for developing measures and yardsticks, standards and criteria in respect of soil conservation schemes.

Statistics of Animal Husbandry Products

Data on livestock and poultry numbers and agricultural implements are collected quinquennially by the Directorate of Economics and Statistics in the Ministry of Food and Agriculture on complete enumeration basis through the primary agencies of the State Governments. The last four livestock censuses have been conducted on improved lines and provide reliable data on the various categories of livestock population and agricultural implements. The last livestock census conducted in 1966 provide data on breed, detailed age composition, calving interval and lactation period also. The scope of this census was extended to collect data on fishing crafts and tackles too.

The Directorate of Marketing and Inspection (DMI) frame estimates of production, utilisation, demand, etc., of livestock products based on the marketing surveys carried out by them. The estimates of production of milk given in DMI Marketing Reports are derived from number of milch animals obtained from the Livestock Census and the average milk yield per milch animal estimated on the basis of percentage of animals that calve annually, calving interval, length of lactation and yield per lactation. As the livestock census figures are available quinquennially and the surveys for finding out the average yield of milk of different categories of milch animals are conducted occasionally, year-to-year changes in milk production have become difficult to assess. Further, the various factors in DMI formula are not correctly estimated.

For putting the data on production of milk on a sound footing, the best possible approach would be a detailed sample survey extending over the whole year in which the milk yields of animals in selected households should be recorded by actual weighment. Such a sampling technique has been evolved by the IARS. When these surveys are conducted extensively in different States, it should be possible to provide reliable estimates of annual milk production in the country.

The output of milk products, *e.g.*, ghee, butter, dahi, cream, khoya, etc., is estimated by making use of appropriate conversion ratios, estimates of proportions of milk converted into different products and the ratios of milk to such products. Scientific surveys on regional basis are necessary to obtain reliable estimates of the various proportions and conversion ratios.

Data on meat, hides and skins, wool and hair, bones, eggs, etc., contained in DMI Marketing Reports also suffer from similar defects and it is necessary to conduct sample surveys on the lines developed by IARS, to enhance their reliability.

Agricultural Research Statistics

Evolution of proper methodology for collection of new types of data needed for planning agricultural development and for assessment of such plans involves considerable research and pilot studies. The IARS is engaged in such research investigations.

In order to assess the impact of milk supply schemes on rural areas, it is necessary to collect information on quantity and composition of feed and fodder given to animals, changes in number of milch animals, cropping pattern, assets and equipment connected with dairying, receptivity to adoption of improved animal husbandry practices, consumption and utilisation of milk both in producer and non-producer households, economic status of the producers, etc. The survey for collection of data on these aspects in an area will have to be carried out periodically starting with a bench mark survey prior to the commencement of rural milk collection.

Sample surveys on fertilizer and other manuring practices are being conducted in a few selected districts spread over different States, to collect data on consumption of different fertilizers and manures, area benefited, rates of application, types of farmers using them, associated cultural practices, etc. It is necessary to conduct these surveys in more districts and to repeat them in the same districts after a lapse of five years or so in order to study the change in the extent and pattern of fertilizer consumption.

Incidence of pests and diseases causes appreciable loss of crops. A pilot sample survey was initiated on rice crop in Cuttack district in 1959 to collect data on incidence of pests and diseases. This was extended to Thanjavur and West Godavari in 1962 and 1963. The survey on wheat and maize crops has recently been started in Aligarh. Based on the experience of these pilot surveys, an all-India survey on important crops needs to be planned.

Cost of cultivation surveys have been found useful for guiding agricultural price policy as well as planning. Such surveys on commercial crops like cotton, sugarcane, oilseeds have been conducted. Pilot surveys have been undertaken on arecanut and coconut. A number of important crops, however, remain to be covered in important producing regions. There is also a need to repeat the surveys on crops covered earlier to bring the data up date.

Large-scale random sampling enquiries for estimating cost of pr
tion of milk in urban and rural areas in Delhi (1953-55), Madras (19

and Calcutta (1960-62) have been carried out, and a satisfactory sampling technique has been developed. A study for estimating the availability and cost of milk production in the areas around Bangalore city has been taken up. It would be useful if such enquiries are carried out in more centres, specially around milk supply schemes.

Conclusion

The process of improvement of agricultural statistics, like that in statistics of other sectors, is a continuing one. The planning for agricultural development is often handicapped for want of complete and reliable data. Certain gaps in the existing data are noticed ; the coverage is found inadequate, quality defective, comparability lacking and time-lag too wide. Further, as a development pragramme progresses, fresh data are required for assessment and evaluation of the programme and for formulation of measures to tide over operational difficulties. With the growth of economy as a result of development measures, new problem of formulation and execution of projects for a more intensified and diversified development, crop up. Further improvement in the range, quality and content of data is called for ; some revision in the concepts and definitions and classification adopted, becomes necessary to meet new situations ; new types of data with regard to both old and new enterprises are required. Timeliness in availability of data for planning purposes assumes added importance as the pace of development is accelerated. Thus, in the process of improvement in agricultural statistics every step is important ; different measures for improvement have to follow a definite sequence. Postponement of improvements needed in a particular period not only jeopardises further advancement with regard to the data needed for planning, but also impedes the very process of planning. The improvements in agricultural statistics listed in this note need immediate attention so that the basic material for formulation of the projects for agricultural development in the manner and according to the technique advocated under the successive Five-Year-Plans, becomes available, and the necessary background is created for further improvements in the statistical information needed for planning for more intensified and diversified development of the agricultural economy of the country.

INVERSE PROBABILITY AND CONFIDENCE INTERVALS

By

V. S. HUZURBAZAR

(*University of Poona*)

1. Introduction. It is an interesting fact that in many problems of statistical estimation the results given by the theory of inverse probability (as modified by Jeffreys) are indistinguishable from those given by the methods of the direct theory such as fiducial probability or confidence intervals. The derivation of some of the inverse distributions by Jeffreys[8] arouses one's curiosity. It seems that when this agreement is noticed there are usually sufficient statistics for parameters in the distribution. One aspect of this agreement has been studied by Huzurbazar[1].

The sufficient statistics cases where agreement is noticed between the direct approach and the inverse probability approach belong to the regular type, *i.e.*, when the range of a distribution does not depend on the parameter to be estimated. It is shown in this paper that for distributions admitting a sufficient statistic in the non-regular case (*i.e.*, when the range of a distribution depends on the parameter), the method of inverse probability leads to the same results arrived at by the method of confidence intervals.

For convenience the following notation in Jeffreys' probability logic is used below :

$P(q/p)$ is the probability of a proposition q on data p.

$P(dx/\alpha H)$ denotes the probability that x lies in a particular range dx, given the previous information H which also includes the known mathematical form of the probability law of x, and α is a particular value of the parameter α occurring in the law.

$P(d\alpha/H)$ is the prior probability differential of the parameter α when it is unknown.

θ will be used to represent the observational data, so that $P(d\alpha/\theta H)$ is the posterior probability differential of α.

2. The general form of distributions admitting a sufficient statistic for a parameter in the non-regular case.

Pitman[9] has given the general form of distributions admitting a sufficient statistic for a parameter α when the range of the distribution depends on α :

$$f(x\,;\,\alpha)=\frac{g(x)}{h(\alpha)}\qquad[a(\alpha)\leqslant x\leqslant b(\alpha)], \qquad \ldots (1)$$

where $g(x)$ is a function of x only, $h(\alpha)$ is a function of α only ; $a(\alpha)$ and $b(\alpha)$ are monotonic in α in opposite senses. In special cases, one of the extremities a or b may be fixed. A rigorous and extensive treatment of Pitman's result on non-regular cases has been given by Huzurbazar[5].

Let $a^{-1}(\alpha)$ and $b^{-1}(\alpha)$ be the functions inverse to $a(\alpha)$ and $b(\alpha)$ respectively. Let $X_{(1)}$ and $X_{(n)}$ be respectively the smallest and greatest members in a random sample of n independent observations from the distribution (1). Let $\hat{\alpha}$ be equal to the smaller of $a^{-1}(X_{(1)})$ and $b^{-1}(X_{(n)})$, if $a(\alpha)$ is monotone increasing and $b(\alpha)$ is monotone decreasing. Let $\hat{\alpha}$ be equal to the greater of $a^{-1}(X_{(1)})$ and $b^{-1}(X_{(n)})$, if $a(\alpha)$ is monotone decreasing and $b(\alpha)$ is monotone increasing. Then $\hat{\alpha}$ is a sufficient statistic for α. As shown by Huzurbazar[2], $\hat{\alpha}$ is also the maximum likelihood estimator of α.

Two well-known distributions included in the family (1) are

(*i*) The rectangular distribution with range α.

$$f(x\ ;\ \alpha) = \frac{1}{\alpha} \qquad \left(-\frac{1}{2}\alpha \leqslant x \leqslant \frac{1}{2}\alpha\right) \qquad \ldots(2)$$

or

$$f(x\ ;\ \alpha) = \frac{1}{\alpha} \qquad (0 \leqslant x \leqslant \alpha) \qquad \ldots(3)$$

(*ii*) The exponential distribution

$$f(x\ ;\ \alpha) = e^{-(x-\alpha)} \quad (\alpha \leqslant x < \infty). \qquad \ldots(4)$$

3. **Confidence intervals for α.**

Let $\epsilon (0 < \epsilon < 1)$ be the confidence coefficient. It has been shown by Huzurbazar[3] that the confidence limits for α with confidence coefficient ϵ are

$$h^{-1}[(1-\epsilon)^{-\frac{1}{n}}\ h(\hat{\alpha})] \text{ and } \hat{\alpha}, \qquad \ldots(5)$$

if $a(\alpha)$ is monotone increasing and $b(\alpha)$ is monotone decreasing ; and

$$\hat{\alpha} \text{ and } h^{-1}\ [(1-\epsilon)^{-\frac{1}{n}}\ h(\hat{\alpha})], \qquad \ldots(6)$$

if $a(\alpha)$ is monotone decreasing and $b(\alpha)$ is monotone increasing. In (5) and (6), n denotes the sample-size.

Confidence intervals for the parameter α of the rectangular distribution and the exponential distribution are obtained by putting $h(\alpha) = \frac{1}{\alpha}$ in (6) and $h(\alpha) = e^{-\alpha}$ in (5) respectively.

4. Inverse Probability.

Huzurbazar[4] has developed a theory of new invariants of distributions admitting sufficient statistics for parameters. An account of these invariants is given by Huzurbazar[7]. A brief account of Huzurbazar's invariants in regular cases has been also given by Jeffreys [8, p. 189] and Huzurbazar[6].

For distributions of the type (1), Huzurbazar [4 and 7] has given the interesting invariance rule : "Take the prior probability of *dh* to be proportional to *dh/h*". This rule is found to be satisfactory for practical distributions, and it has the advantage that it is invariant for all one-to-one transformations of the variate and the parameter. The rule therefore ensures consistency in prior probability statements. In particular, the rule $P(dh/H) \propto dh/h$ gives the prior probability form $d\alpha/\alpha$ for the parameter of the rectangular distribution. This is the appropriate prior probability form used by Jeffreys in this case, since α is a scale parameter. The invariance rule leads to the prior probability form $P(d\alpha/H) \propto d\alpha$ for the exponential distribution. This is also quite satisfactory, following Jeffreys, since α is a location parameter in this case.

First suppose that $a(\alpha)$ is monotone increasing and $b(\alpha)$ is monotone decreasing, so that the confidence limits for α with confidence coefficient ϵ are given by the (5).

Let $$P(dh/H) \propto \frac{dh}{h}. \qquad \ldots(7)$$

If $x_1, x_2 \ldots\ldots x_n$ is a random sample of size n from distribution (1), we have

$$P(dx_1 dx_2 \ldots\ldots dx_n/hH) = \frac{g(x_1)g(x_2)\ldots g(x_n)}{[h(\alpha)]^n}\, dx_1 dx_2 \ldots dx_n \qquad \ldots(8)$$

Then by the principle of inverse probability,

$$P(dh/\theta H) \propto P(dh/H)\ P(dx_1 dx_2\ .\ dx_n/hH)$$
$$\propto \frac{dh}{h^{n+1}}$$

Let $$P(dh/\theta H) = \frac{\lambda}{h^{n+1}}\, dh, \qquad \ldots(9)$$

where λ is a constant.

In this case $h(\alpha)$ is a monotone decreasing function of α. Moreover, given the sample, the range of possible value of α is given by

$$\alpha \leqslant \hat{\alpha}, \qquad \ldots(10)$$

where $\hat{\alpha}$ is the maximum likelihood estimator of α.

Now (10) is equivalent to

$$h(\alpha) \geqslant h\hat{\alpha} \qquad \ldots(11)$$

so that the range of possible values of $h(\alpha)$, given the sample is $[h(\hat{\alpha}), \infty]$.

The constant λ in (9) is determined by

$$\lambda \int_{h(\hat{\alpha})}^{\infty} \frac{dh}{h^{n+1}} = 1, \qquad \ldots(12)$$

which gives

$$\lambda = n[h(\hat{\alpha})]^n. \qquad \ldots(13)$$

Hence the posterior probability distribution of $h(\alpha)$ is given by

$$P(dh/\theta H) = \frac{n[h(\hat{\alpha})]^n}{h^{+1}} dh \quad [h(\hat{\alpha}) \leqslant h < \infty] \qquad \ldots(14)$$

Now consider the inequality

$$h(\hat{\alpha}) \leq h(\alpha) \leqslant (1-\epsilon)^{-1/n} \, h(\hat{\alpha}) \qquad \ldots(15)$$

Using (14), the probability of inequality (15) is

$$\int_{h(\hat{\alpha})}^{(1-\varepsilon)^{-1/n} h(\hat{\alpha})} \frac{nh[(\hat{\alpha})]^n}{h^{n+1}} \, dh = \epsilon \qquad \ldots(16)$$

But since $h(\alpha)$ is monotone decreasing, the inequality (15) is equivalent to the inequality

$$\hat{\alpha} \geqslant \alpha \geqslant h^{-1}[(1-\epsilon)^{-\frac{1}{n}} \, h(\hat{\alpha})],$$

i.e.

$$h^{-1}[(1-\epsilon)^{-\frac{1}{n}} h(\hat{\alpha})] \leqslant \alpha \leqslant \hat{\alpha} \qquad \ldots(17)$$

Hence by the method of inverse probability, the probability of the inequality (17) is ϵ, which is also the confidence coefficient of the same inequality. The proof when $a(\alpha)$ is decreasing and $b(\alpha)$ is increasing is on similar lines. Hence the method of inverse probability leads to the same results as the method of confidence intervals.

REFERENCES

1. Huzurbazar, V.S. (1948) : "Inverse probability and sufficient statistics". Proc. Camb. Phil. Soc., **45**, 225.

2. Huzurbazar, V.S. (1948) ; "The likelihood equation, consistency and the maxima of the likelihood function". Ann. Eugen ; Lond ; **14**, 185.

3. Huzurbazar, V.S. (1955) : "Confidence intervals for the parameter of a distribution admitting a sufficient statistic when the range depends on the parameter." J.R.S.S. (B), **17**, 86.

4. Huzurbazar, V.S. (1960) : *Invariance theory of prior probabilities* (Unpublished ; Adams Prize Essay, University of Cambridge, England).

5. Huzurbazar, V.S. (1964) : "The general forms of distribution admitting sufficient statistics for parameters in nonregular cases" (*Properties of sufficient statistics*: Report submitted to the National Science Foundation of U.S.A.).

6. Huzurbazar, V.S. (1966) : "Some invariants of some discrete distributions admitting sufficient statistics for parameters". *Classical and contagious discrete distributions*, Proceedings of the International Symposium (Montreal 1963), 231. Also reprinted in Sankhya, A. **28** (1966), 215.

7. Huzurbazar, V.S. (1967) : "Invariants of probability distributions". Presidential address, Section of Statistics 54th Session of the Indian Science Congress, Hyderabad.

8. Jeffreys, H. (1961) : *Theory of Probability*, 3rd ed. ; Oxford, Clarendon Press.

9. Pitman, E.J.G. (1936) : "Sufficient statistics and intrinsic accuracy". Proc. Camb. Phil. Soc. **32**, 567.

MODIFIED GOODNESS OF FIT TESTS FOR HYPOTHESIS OF *MARKOV CHAINS*

By
P.V. KRISHNA IYER
Punjab University
and
P. SAMARASIMHUDU
Defence Science Laboratory

Abstract. The paper gives three theorems useful in the evaluation of cumulants of Psi-squared statistics for Markov chains. Using these theorems, the variance and covariance for transitions of different orders and the first two asymptotic cumulants of ψ_1^2, ψ_2^2 and ψ_3^2 have been evaluated. The expected value andvariance of ψ_t^2 for a random sequence have also been given. The distribution of ψ_t^2 is approximated to the χ^2 form by assuming χ^2 equal to $A\psi^2/B$ with A^2/B degrees of freedom where A and $2B$ are the expected value and variance of ψ^2.

1. Introduction. In many scientific investigations we are faced with the problem of testing the hypothesis whether a given sequence of observations, in which the variables take a finite number of values, say a, b, is random (independent and unordered) or whether there is some dependence between the observations. One form of dependence for which a good deal of work has been done during the past fifteen years is the Markovian dependence of different orders. In first order Markovian chains there is dependence between successive observations only. The chain in this case is defined by the conditional probability matrix (p_{ij}), where p_{ij} is the conditional probability that i will be followed by j. If there is dependence of the second order, the probability matrix is defined by a $k^2 \times k^2$ matrix. The conditional probability in this case is defined by p_{ijk} which represents the probability that ij is followed by jk. The probability that ij is followed by lm is zero provided $j \neq l$. It may be noted that the probability $P(ij)$ that i will be followed by j is $P_i \, p_{ij}$, where P_i is the asymptotic probability for i in the sequence. Similarly $P_i p_{ij} p_{ijk} \simeq P(ijk)$.

Now to test the hypothesis of Markovian dependence of a given sequence of observations, the best procedure is to use the likelihood ratio. Bartlett (1), Billingsley (3), Good (4) and Goodman (5) and a number of others have examined this problem in great detail and have shown that the likelihood ratio can be approximated to the Psi-squared goodness of fit criterion which is algebraically similar to Pearson's goodness of fit criterion. This test is defined by

$$\psi_t^2 = \sum_u \frac{(n_{u_t} - m_{u_t})^2}{m_{u_t}}$$

where n_{u_t} represents the frequency that $(t+1)$ consecutive observations will take the values $i.j. .. m$ in the order specified, extends to all possible combinations of $(t+1)$ observations from $a, b, \ldots; k$; and m_{u_t} is the expected number of $(t+1)$ observations defined by u_t. Thus

$$\psi_1^2 = \sum_{i=1}^{n} \frac{(n_i - m_i)^2}{m_i}$$

$$\psi_2^2 = \sum_{i,j} \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

$$\psi_3^2 = \sum_{i,j,k} \frac{(n_{ijk} - m_{ijk})^2}{m_{ijk}}$$

The distribution of the above statistics is complicated. It cannot be obtained from the usual χ^2-distribution which has been tabulated extensively. In fact, $\psi_1^2, \psi_2^2, \ldots, \psi_t^2$ is distributed at the weighted sum of a number of X^2 with the degrees of freedom indicated in the suffix.

$$\begin{aligned}
\psi_1^2 &= \chi_{s-1}^2 \\
\psi_2^2 &= \chi_{(s-1)^2}^2 + 2\chi_{s-1}^2 \\
\psi_3^2 &= \chi_{s(s-1)^2}^2 + 2\chi_{(s-1)^2}^2 + 3\chi_{s-1}^2 \\
\ldots \quad & \quad \ldots \qquad \ldots \\
\psi_t^2 &= \chi_{s(t-2)\,(s-1)^2}^2 + 2\chi_{s(t-3)\,(s-1)^2}^2 + \ldots \\
&\qquad + (r-1)\ \chi_{s^{(t-r)}\,(s-1)^2}^2 + \ldots + t\chi_{(s-1)}^2
\end{aligned}$$

Following Patankar (7) these distributions can be approximated to modified χ^2 distribution where the modified χ^2 is obtained from $A\psi^2/B$ with A^2/B degrees of freedom, where A and $2B$ are the expected value and variance of ψ^2. Therefore, if we know the expected value and variance of ψ_t^2 the Psi-square test can be used as a test of hypothesis for testing Markov sequences of different orders. Bhat (2) has given the variance and the expected value of ψ_t^2, but the form in which this has been given is not in a form that can be readily used for purposes of calculations. In this paper we present three theorems which can be used for obtaining the various summations arising in the evaluation of the variances for ψ_1^2, ψ_2^2 and ψ_3^2 statistics. They appear to be fundamental for investigations in Psi-squared statistics. Using them, variance and covariance for transitions of different orders and the first two cumulants of the statistics ψ_1^2, ψ_2^2 and ψ_3^2 have been evaluated. The expected value and variance of ψ_t^2 for a random sequence have also been given.

It is obvious from the above considerations that the Psi-squared test can be reduced to the well known Pearson's χ^2-test wherein $\chi^2 = A\psi^2/B$ with A^2/B degrees of freedom. Therefore, if the expected value of ψ^2 and its variance are known, the existing χ^2-table can be used to test the hypothesis of Markovian dependence.

2. Some Useful Theorems.

For a Markov chain of n observations defined by the conditional probability matrix (p_{ij}), we prove the following theorems :

Theorem 1

$$\sum_{s=1}^{n-1}\sum_{i=a}^{k}\left(1-\frac{s}{n}\right)p_{ii}\left(P_{ii}^{(s)}-p_i\right)+\sum_{s=1}^{n-1}\sum_{\substack{i,\,j\\ i\neq j}}\left(1-\frac{s}{n}\right)p_{ij}\left(p_{ji}^{(s)}-P_i\right)$$

$$=\sum_{s=1}^{n-1}\sum_{i}\left(1-\frac{s}{n}\right)\left(p_{ii}^{(s+1)}-P_i\right), \qquad \text{...(2.1.1)}$$

where i, j represent the various states, $p_{ji}^{(s)}$ the conditional probability of $(s+1)$th observation to be in state i when the first observation is in state j (the intermediate observations being in any state) and P_i the asymptotic probability for an observation to be in state i.

Proof. For a particular value of i, the left hand side becomes

$$\sum_{s=1}^{n-1}\left(1-\frac{s}{n}\right)p_{ii}\left(p_{ii}^{(s)}-P_i\right)+\sum_{s=1}^{n-1}\sum_{\substack{j=a\\ j\neq i}}^{k}\left(1-\frac{s}{n}\right)p_{ij}\left(p_{ji}^{(s)}-P_i\right)$$

$$=\sum_{s=1}^{n-1}\sum_{j=a}^{k}\left(1-\frac{s}{n}\right)p_{ij}\left(p_{ji}^{(s)}-P_i\right)$$

$$=\sum_{s}\left(1-\frac{s}{n}\right)\Big[p_{ia}\left(p_{ai}^{(s)}-P_i\right)+p_{ib}\left(p_{bi}^{(s)}-P_i\right)+\dots\dots\dots$$

$$\dots\dots+p_{ik}\left(p_{ki}^{(s)}-P_i\right)\Big]. \qquad \text{...(2.1.2)}$$

Denote $\quad p_{ia}\;p_{ai}^{(s)}+p_{ib}\;p_{bi}^{(s)}\;\dots\dots\dots+p_{ik}\;p_{ki}^{(s)}$ by θ.

Then $\quad P_1\theta=P_1\Big[p_{ia}\;p_{ai}^{(s)}+p_{ib}\;p_{ib}^{(s)}\;\dots\dots+p_{ik}p_{ki}^{(s)}\Big] \qquad \text{...(2.1.3)}$

The first term in the R.H.S. of (2.1.3) gives the joint probability for the first observation to be in state i, followed by the second observation in state a and the $(s+2)$th observation in state i. Similar definitions can be given for the other terms of the R.H.S. of (2.1.3) since $a, b, c, \dots\dots, k$ occur immediately after i in these terms and are followed by i in $(s+2)$th place

$$P_i\theta=P_i\;p_{ii}^{(s+1)}$$

$$\therefore \qquad \theta = p_{ii}^{(s+1)}\,.$$

Hence (2.1.2) reduces to $\sum_{s=1}^{n-1}\left(1-\frac{s}{n}\right)\left(p_{ii}^{(s+1)}-P_i\right)$.

Hence the result.

Theorem II

$$p_{ij}\,p_{rm}\left[P_i\sum_{s=1}^{n-1}\left(1-\frac{s}{n}\right)\left(p_{jr}^{(s)}-P_r\right)+P_r\sum_{s=1}^{n-1}\left(1-\frac{s}{n}\right)\left(p_{mi}^{(s)}-P_i\right)\right]^2$$

$$\approx 4P_iP_r\,p_{ij}\,p_{rm}\sum_{s_1,s_2=1}^{n-1}\left(1-\frac{s_1}{n}\right)\left(1-\frac{s_2}{n}\right)\left(p_{jr}^{(s_i)}-P_r\right)\times\left(p_{mi}^{(s_2)}-P_i\right)$$

Proof. The L.H.S. is equivalent to

$$p_{ij}p_{rm}\left[P_i\left\{\sum_s\left(1-\frac{s}{n}\right)\left(p_{jr}^{(s)}-P_r\right)\right\}-P_r\left\{\sum_s\left(1-\frac{s}{n}\right)\left(p_{mi}^{(s)}-P_i\right)\right\}\right]^2+4p_{ij}p_{rm}\sum_{s_1,\,s_2}P_iP_r\left(1-\frac{s_1}{n}\right)\left(1-\frac{s_2}{n}\right)\left(p_{jr}^{(s_1)}-P_r\right)\left(p_{mi}^{(s_2)}-P_i\right)$$

The first term is very small as compared to the second term. This can be established as follows :

$$\underset{s\to\infty}{\mathrm{Lt}}\; p_{jr}^{(s)} \to P_r,\quad \underset{s\to\infty}{\mathrm{Lt}}\; p_{mi}^{(s)} \to P_i$$

or

$$\underset{s\to\infty}{\mathrm{Lt}}\;\frac{p_{jr}^{(s)}}{P_r} = \underset{s\to\infty}{\mathrm{Lt}}\;\frac{p_{mi}^{(s)}}{P_i} = 1$$

or

$$\underset{s\to\infty}{\mathrm{Lt}}\;\frac{p_{jr}^{(s)}-P_r}{P_r} = \underset{s\to\infty}{\mathrm{Lt}}\;\frac{p_{mi}^{(s)}-P_i}{P_i} = 0$$

For s finite, let

$$\frac{p_{jr}^{(s)}-P_r}{P_r} = \alpha_{r_s}$$

and

$$\frac{p_{mi}^{(s)}-P_i}{P_i} = \beta_{r_s}$$

The first term becomes

$$p_{ij}\,p_{rm}\ \Sigma\left[P_i\,P_r\left(1-\frac{s}{n}\right)\alpha_{r_s}\ -P_rP_i\left(1-\frac{s}{n}\right)\beta_{r_s}\ \right]^2$$

$$=p_{ij}p_{rm}\mathrm{P}_i^2\ \mathrm{P}_r^2\ \left[\Sigma\left(1-\frac{s}{n}\right)\left(\alpha_{r_s}-\beta_{r_s}\right)\right]^2$$

Let

$$\alpha_{r_s}=\alpha+\epsilon_1$$

$$\beta_{r_s}=\alpha+\epsilon_2$$

$$\epsilon=\epsilon_1+\epsilon_2$$

where ϵ_1 and ϵ_2 are very small quantities, then the first term reduces to

$$p_{ij}p_{rm}\ \mathrm{P}_i^2\ \mathrm{P}_r^2\ \left[\ \Sigma\left(1-\frac{s}{n}\right)(\epsilon_1-\epsilon_2)\ \right]^2$$

and the second term becomes

$$4p_{ij}p_{rm}\ \mathrm{P}_i^2\ \mathrm{P}_r^2\ \left[\sum_{s_1,s_2}\left(1-\frac{s_1}{n}\right)\left(1-\frac{s_2}{n}\right)\alpha_{r_s}\ \beta_{r_s}\ \right]$$

$$=4p_{ij}p_{rm}\ \mathrm{P}_i^2\ \mathrm{P}_r^2\ \sum_{s_1,s_2}\left(1-\frac{s_1}{n}\right)\left(1-\frac{s_2}{n}\right)(\alpha^2+\alpha\epsilon+\epsilon_1\epsilon_2).$$

We see that the first term is very small as compared to second term. This is true for all values of *i*, *j*, *r* and *m*. Hence the result.

Theorem III

$$\sum_{s_1,s_2=1}^{n-1}\left(1-\frac{s_1}{n}\right)\left(1-\frac{s_2}{n}\right)\left[\sum_i p_{ii}^2\left(p_{ii}^{(s_1)}-\mathrm{P}_i\right)\left(p_{ii}^{(s_2)}-\mathrm{P}_i\right)\right.$$

$$+\sum_{\substack{i,j\\ i\neq j}}p_{ij}^2\left(p_{ji}^{(s_1)}-\mathrm{P}_i\right)\left(p_{ji}^{(s_2)}-\mathrm{P}_i\right)+\sum_{\substack{i,j\\ i\neq j}}\mathrm{P}_{ii}\,\mathrm{P}_{ij}\left(p_{ii}^{(s_2)}-\mathrm{P}_i\right)\left(p_{ji}^{(s_2)}-\mathrm{P}_i\right)$$

$$+\sum_{\substack{i,j\\ i\neq j}}p_{ii}\,p_{ji}\left(p_{ij}^{(s_1)}-\mathrm{P}_j\right)\left(p_{ii}^{(s_2)}-\mathrm{P}_i\right)+\sum_{\substack{i,j\\ i\neq j}}p_{ij}\,p_{ji}\left(p_{jj}^{(s_1)}-\mathrm{P}_j\right)\left(p_{ii}^{(s_2)}-\mathrm{P}_i\right)$$

$$+\sum_{\substack{i,j\\ i\neq j}}p_{ii}\,p_{ji}\left(p_{ij}^{(s_1)}-\mathrm{P}_j\right)\left(p_{ji}^{(s_2)}-\mathrm{P}_i\right)+\sum_{\substack{i,j,m\\ i\neq j\neq m}}p_{ii}\,p_{jm}\left(p_{ij}^{(s_1)}-\mathrm{P}_j\right)\left(p_{mi}^{(s_2)}-\mathrm{P}_i\right)$$

$$\left.+\sum_{\substack{i,j,m\\ i\neq j\neq m}}p_{ij}\,p_{jm}\left(p_{jj}^{(s_1)}-\mathrm{P}_j\right)\left(p_{mi}^{(s_2)}-\mathrm{P}_i\right)+\sum_{\substack{i,j,m,n\\ i\neq j\neq m\neq n}}\mathrm{P}_{ij}\left(p_{jm}^{(s_1)}-\mathrm{P}_m\right)\left(p_{ni}^{(s_2)}-\mathrm{P}_i\right)\right]$$

$$=\sum_{s_1,\ s_2=1}^{n-1}\ \sum_{i=a}^{k}\left(1-\frac{s_1}{n}\right)\left(1-\frac{s_2}{n}\right)\left(p_{ii}^{(s_1+s_2+2)}-\mathrm{P}_i\right),$$

where *i*, *j*, *m*, *n* represent *a*, *b*, *c*……, *k* with the relevant restrictions.

Proof. All the probabilities on the L.H.S. of the identity arise from the squares of variance and covariance terms of the X's is in the $k \times k$ contingency table shown below where X_{ij} represents the number of transitions from i to j.

		Succeeding states				
		a	b	c	d	k
PRECEDING STATES	a	X_{aa}	X_{ab}	X_{ac}	X_{ad}	X_{ak}
	b	X_{ba}	X_{bb}	X_{bc}	X_{bd}	X_{bk}
	c	X_{ca}	X_{cb}	X_{cc}	X_{cd}	X_{ck}
	d	X_{da}	X_{db}	X_{dc}	X_{dd}	X_{dk}
	k	X_{La}	X_{kb}	X_{kc}	X_{kd}	X_{kk}

The total number of combinations arising in L.H.S. is k^4. Each of the $X_{ij}(i, j=a, b, \ldots, k)$ is associated with all the k^2 terms. Taking a particular transition X_{aa} and associating it with all the other (k^2-1) transitions and itself, *i e.*, (the contribution from its variance) we get the following results.

First associating its with the first row *i.e.*, (X_{aa}, X_{aa}), (X_{aa}, X_{ab}), (X_{aa}, X_{ac}), ..., (X_{aa}, X_{ak}) the sum of the probabilities arising from squares of variances and covariances in the expression (5·5) mentioned in section 5, we obtain

$$\begin{aligned}
& p_{aa}\left(p_{aa}^{(s_1)} - P_a\right) p_{aa}\left(p_{aa}^{(s_2)} - P_a\right) \\
+ \quad & p_{aa}\left(p_{aa}^{(s_1)} - P_a\right) p_{ab}\left(p_{ba}^{(s_2)} - P_a\right) \\
+ \quad & p_{aa}\left(p_{aa}^{(s_1)} - P_a\right) p_{ac}\left(p_{ca}^{(s_2)} - P_a\right) \\
+ \quad & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
+ \quad & p_{aa}\left(p_{aa}^{(s_1)} - P_a\right) p_{ak}\left(p_{ka}^{(s_2)} - P_a\right) \\
= p_{aa} & \left(p_{aa}^{(s_1)} - P_a\right)\left(p_{aa}^{(s_2+1)} - P_a\right)
\end{aligned}$$

following the arguments of Theorem I and using the property that $\sum\limits_{j=a}^{k} p_{ij}^{(s)} = 1$ for all values of s. When we associate X_{aa} with the second row, the sum of their probabilities

$$p_{aa}\left(p_{ab}^{(s_1)} - P_b\right) p_{ba}\left(p_{aa}^{(s_2)} - P_a\right)$$

$$\begin{aligned}
+ \quad & p_{aa}\left(p_{ab}^{(s_1)} - P_b \right) p_{bb} \left(p_{ba}^{(s_2)} - P_a \right) \\
+ \quad & p_{aa}\left(p_{ab}^{(s_1)} - P_b \right) p_{bc} \left(p_{ca}^{(s_2)} - P_a \right) \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
+ \quad & p_{aa}\left(p_{ab}^{(s_1)} - P_b \right) p_{bk} \left(p_{ka}^{(s_2)} - P_a \right) \\
= \quad & p_{aa}\left(p_{ab}^{(s_1)} - P_b \right)\left(p_{ba}^{(s_2+1)} - P_a \right)
\end{aligned}$$

Similarly when we associate X_{aa} with the third row we get

$$p_{aa}\left(p_{ac}^{(s_1)} - P_e \right)\left(p_{ca}^{(s_2+1)} - P_a \right)$$

and lastly with the kth, we get

$$p_{aa}\left(p_{ak}^{(s_1)} - P_k \right)\left(p_{ka}^{(s_2+1)} - P_a \right)$$

Now taking the total contribution of X_{aa}, *i.e.*

$$\begin{aligned}
& p_{aa}\left(p_{aa}^{(s_1)} - P_a \right)\left(p_{aa}^{(s_2+1)} - P_a \right) \\
+ \quad & p_{aa}\left(p_{ab}^{(s_1)} - P_b \right)\left(p_{ba}^{(s_2+1)} - P_a \right) \\
+ \quad & p_{aa}\left(p_{ac}^{(s_1)} - P_c \right)\left(p_{ca}^{(s_2+1)} - P_a \right) \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
+ \quad & p_{aa}\left(p_{ak}^{(s_1)} - P_k \right)\left(p_{ka}^{(s_2+1)} - P_a \right) \\
= \quad & P_{aa}\left(p_{aa}^{(s_1+s_2+1)} - P_a \right)
\end{aligned}$$

Next associating X_{ab} with the first row, we get

$$\begin{aligned}
& p_{ab}\left(p_{ba}^{(s_1)} - P_a \right) p_{aa}\left(p_{aa}^{(s_2)} - P_a \right) \\
+ \quad & p_{ab}\left(p_{ba}^{(s_1)} - P_a \right) p_{ab}\left(p_{ba}^{(s_2)} - P_a \right) \\
+ \quad & p_{ab}\left(p_{ba}^{(s_1)} - P_a \right) p_{ac}\left(p_{ca}^{(s_2)} - P_a \right) \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
+ \quad & p_{ab}\left(p_{ba}^{(s_1)} - P_a \right) p_{ak}\left(p_{ka}^{(s_2)} - P_a \right) \\
= \quad & p_{ab}\left(p_{ba}^{(s_1)} - P_a \right)\left(p_{aa}^{(s_2+1)} - P_a \right)
\end{aligned}$$

Associating X_{ab} with the second row, we get

$$p_{ba}\left(p_{bb}^{(s_1)}-P_b\right)\, p_{ba}\left(p_{aa}^{(s_2)}-P_a\right)$$

$$+\quad p_{ab}\left(p_{bb}^{(s_1)}-P_b\right)\, p_{bb}\left(p_{ba}^{(s_2)}-P_a\right)$$

$$+\quad p_{ab}\left(p_{bb}^{(s_1)}-P_b\right)\, p_{bc}\left(p_{ca}^{(s_2)}-P_a\right)$$

....

$$+\quad p_{ab}\left(p_{bb}^{(s_1)}-P_b\right)\, p_{bk}\left(p_{ka}^{(s_2)}-P_a\right)$$

$$=\quad p_{ab}\left(p_{bb}^{(s_1)}-P_b\right)\left(p_{ba}^{(s_2+1)}-P_a\right)$$

Similarly considering it in relation to other rows and adding all of them, we get

$$p_{ab}\left(p_{ba}^{(s_1)}-P_a\right)\left(p_{aa}^{(s_2+1)}-P_a\right)$$

$$+\quad p_{ab}\left(p_{bb}^{(s_1)}-P_b\right)\left(p_{ba}^{(s_2+1)}-P_a\right)$$

$$+\quad p_{ab}\left(p_{bc}^{(s_1)}-P_c\right)\left(p_{ca}^{(s_2+1)}-P_a\right)$$

....

$$+\quad p_{ab}\left(p_{bk}^{(s_1)}-P_k\right)\left(p_{ka}^{(s_2+1)}-P_a\right)$$

$$=\quad p_{ab}\left(p_{ba}^{(s_1+s_2+1)}-P_a\right)$$

Following the same procedure, we get the contribution of X_{ac} as

$$p_{ac}\left(p_{ca}^{(s_1+s_2+1)}-P_a\right)$$

and lastly the contribution of X_{ak} as

$${}_{ak}\left(p_{ka}^{(s_1+s_2+1)}-P_a\right).$$

Multiplying the contributions of X_{aa}, X_{ab}, $X_{ac}, \dots X_{ak}$ by P_a and adding, we get

$$P_a p_{aa}\left(p_{aa}^{(s_1+s_2+1)}-P_a\right)$$

$$+\quad P_a\, p_{ab}\left(p_{ba}^{(s_1+s_2+1)}-P_a\right)$$

$$+P_a p_{ac}\left(p_{ca}^{(s_1+s_2+1)}-P_a\right)$$

....

$$+P_a\, d_{ak}\left(p_{ka}^{(s_1+s_2+1)} - p_a \right)$$

which is equal to $P_a \left(p_{aa}^{(s_1+s_2+2)} - P_a \right)$

Hence the contribution of the first row is

$$\left(p_{aa}^{(s_1+s_2+2)} - P_a \right)$$

Similarly the contribution of the second row is

$$\left(p_{bb}^{(s_1+s_2+2)} - P_a \right)$$

Therefore the total contribution is

$$\sum_{s_1, s_2=1}^{n-1} \sum_i \left(1 - \frac{s_1}{n}\right)\left(1 - \frac{s_2}{n}\right)\left(p_{ii}^{(s_1+s_2+2)} - P_i \right)$$

3. Cumulants for First Order Chains

(I) Different States

Let X_i be the number of observations in state i in the sequence. Then

$$X_i = x_1, x_2 + \dots + x_n$$

where $x_r \begin{cases} =1 \text{ if } x_r \text{ is in state } i \\ =0 \text{ otherwise.} \end{cases}$

$$\begin{aligned} E(X_i) &= E(x_1 + x_2 + \dots + x_n) \\ &= nE(x_r) \\ &= nP_i \end{aligned}$$

where P_i is the asymptotic probability for an observation to be in state i. (3.1.1)

$$E(X_i^2) = nE(X_r^2) + 2 \sum_{s=1}^{n-1} (n-s)E(x_r, x_{r+s})$$

$E(x_r, x_{r+s})$ = The probability for the rth and $(r+s)$th observations to be in state i, when nothing is known of the observations between rth and $(r+s)$th observations.

$= P_i\, p_{ii}^{(s)}$ where $p_{ii}^{(s)}$ is the conditional probability for $(r+s)$th observation to be in state i when rth observation is in state i.

$$k_2(X_i) = nP_i \left[(1 - P_i) + 2 \sum_{s=1}^{n-1} \left(\frac{n-s}{n} \right)\left(p_{ii}^{(s)} - P_i \right) \right]$$

$$= nP_i \left[Q_i + 2 \sum_{s=1}^{n-1} \left(1 - \frac{s}{n}\right) P_i^{(s)} \right] \qquad \dots(3.1.2)$$

where $$P_{ii}^{(s)} = \left(p_{ii}^{(s)} - P_i\right).$$

$$k_{11}(X_i,\ X_j) = -nP_iP_j + n(P_i\ \alpha'_{ij} + P_j\alpha'_{ji}) \qquad \ldots(3.1.2)$$

where $$\alpha_{ij}' = \sum_{s=1}^{n-1} \left(1 - \frac{s}{n}\right)\left(p_{ij}^{(s)}\right)$$

(II) Transitions like (ii) and (i j) between successive observations

Let X_{ii} be the number of (*ii*) transitions between successive observations in the sequence.

Then $$X_{ii} = x_{12} + x_{23} + \ldots x_{n-1,\ n}$$

where $x_{r,\ r+1}$ represents the transition between r and $(r+1)$th observations and

$$X_{r,r+1} \begin{cases} =1 \text{ if the } r\text{th and } (r+1)\text{th observations are in state } i. \\ =0 \text{ otherwise} \end{cases}$$

$$\begin{aligned} E(X_{ii}) &= (n-1)E(x_{r,\ r+1}) \\ &= (n-1)P_ip_{ii} \\ &\approx nR_{ii} \text{ where } R_{ii} = P_ip_{ii} \end{aligned} \qquad \ldots(3.2.1)$$

Evaluating similarly we get

$$k_2(X_{ii}) \approx nR_{ii}\left[1 - 3R_{ii} + 2p_{ii} + 2p_{ii}\sum_{s=1}^{n-3}\left(1 - \frac{s}{n}\right)P_{ii}^{(s)}\right] \ldots(3.2.2)$$

$$k_1(X_{ij}) \approx nR_{ij} \qquad \text{where } R_{ij} = P_i\, p_{ij} \qquad \ldots(3.2.3)$$

and

$$k_2(X_{ij}) \approx nR_{ij}\left[1 - 3R_{ij} + 2p_{ij}\sum_{s=1}^{n-3}\left(1 - \frac{s}{n}\right)P_{ji}^{(s)}\right] \qquad \ldots(3\ 2.4)$$

$$k_{11}(X_{ii},\ X_{ij}) \approx -3nR_{ii}R_{ij} + np_{ii}p_{jj}(P_i\alpha_{ij} + P_j\alpha_{ji}) \qquad \ldots(3.2.5)$$

where $$\alpha_{ij} = \sum_{s=1}^{n-3}\left(1 - \frac{s}{n}\right)\left(P_{ij}^{(s)}\right).$$

$$k_{11}(X_{ii},\ X_{ij}) \approx -3nR_{ii}R_{ij} + nR_{ii}p_{ij} + np_{ii}p_{ij}(P_i\alpha_{ii} + P_ia_{ji}). \qquad \ldots(3.2.6)$$

$$k_{11}(X_{ii},\ X_{jm}) \approx -3nR_{ii}R_{jm} + np_{ii}(P_i\alpha_{ij} + P_j\alpha_{mi}). \qquad \ldots(3.2.7)$$

$$k_{11}(X_{ij},\ X_{jm}) \approx -3nR_{ij}R_{jm} + nR_{ij}p_{jm} + np_{ij}p_{jm}(P_i\alpha_{jj} + P_j\alpha_{mi}) \qquad \ldots(3.2.8)$$

$$k_{11}(X_{ij},\ X_{lm}) \approx -3nR_{ij}R_{lm} + np_{ij}p_{lm}(P_i\alpha_{jl} + p_l\alpha_{mi}) \qquad \ldots(3.2.9)$$

$$k_{11}(X_{ij},\ X_{ji}) \approx -3nR_{ij}R_{ji} + np_{ij}p_{ji}[P_i(1 + \alpha_{jj}) + P_j(1 + \alpha_{ii})] \qquad \ldots(3.2\ 10)$$

$$k_{11}(X_{ii},\ X_{ji}) \approx -3nR_{ii}R_{ji} + nR_{ji}p_{ii} + np_{ii}p_{ji}(P_i\alpha_{ij} + P_j\alpha_{ii}) \qquad \ldots(3.2.11)$$

The results have been briefly given by us[6].

4. Cumulants for Second Order Chains

Let X_{iii} represent the number of (*iii*) transitions between successive observations. Assume (p_{tjl}), (p_{ij}) and (P_i) represent the various transition probabilities. The asymptotic cumulants for various second order transitions reduce to :

$$k_1(X_{iii}) \approx nP_ip_{ii}P_{iii} \qquad \ldots(4.1)$$

$$k_2(X_{iii}) \approx nP_i p_{ii} p_{iii}[1-5P_i p_{ii} P_{iii} + 2p_{iii} + 2p^2{}_{iii} + 2\sum_{s=1}^{n-5}\left(1-\frac{s}{n}\right)\left(p_{ii}^{(s)} - P_i\right)p_{ii}p_{iii}\Bigg] \qquad \ldots(4.2)$$

$$k_1(X_{iij}) \approx nP_i p_{ii} P_{iij}, \qquad \ldots(4.3)$$

$$k_2(X_{iij}) \approx nP_i p_{ii} p_{iij}\Bigg[1-5P_i p_{ii} p_{iii} + + 2\sum_{s=1}^{n-5}\left(1-\frac{s}{n}\right)\left(p_{ji}^{(s)} - P_i\right)p_{ii}p_{iij}\Bigg] \qquad \ldots(4.4)$$

$$k_1(X_{i,i}) \approx nP_i p_{ij} p_{iji}, \qquad \ldots(4.5)$$

$$k_2(X_{iji}) \approx nP_i p_{ij} p_{iji}[1-5P_i p_{ij} p_{iji} + 2p_{iji} P_{jij} + 2\sum_{s=1}^{n-5}\left(1-\frac{s}{n}\right)\left(p_{ii}^{(s)} - P_i\right)p_{ij}\, P_{iji}\Bigg] \qquad \ldots(4.6)$$

$$k_1(X_{ijj}) \approx nP_i p_{ij} p_{ijj}, \qquad \ldots(4.7)$$

$$k_2(X_{ijj}) \approx nP_i p_{ijj}[1-5P_i p_{ij} p_{ijj} + 2\sum_{s=1}^{n-5}\left(1-\frac{s}{n}\right)\left(p_{ji}^{(s)} -- P_i\right)p_{ij}p_{ijj}]. \qquad \ldots(4.8)$$

$$k_1(X_{ijm}) \approx nP_i p_{jj} p_{ijm}' \qquad \ldots(4.9)$$

$$k_2(X_{ijm}) \approx nP_i p_{ij} p_{ijm}\,[1-5P_i p_{ij} p_{ijm} + 2\sum_{s=1}^{n-5}\left(1-\frac{s}{n}\right)\left(p_{mi}^{(s)} - P_i\right)p_{ij}p_{ijm}]. \qquad \ldots(4.10)$$

5. Cumulants of Psi-square Statistics

The statistics ψ_1^2, ψ_2^2 and ψ_3^2 are defined by

$$\psi_1^2 = \sum_i \frac{[X_i - E(X_i)]^2}{E(X_i)} \qquad \ldots(5.1)$$

$$\psi_2^2 = \sum_{i,j} \frac{[X_{ij} - E(X_{ij})]^2}{E(X_{ij})} \qquad \ldots(5.2)$$

and

$$\psi_3^2 = \sum_{i,j,m} \frac{[X_{ijm} - E(X_{ijm})]^2}{E(X_{ijm})} \qquad \ldots(5.3)$$

where $i, j, m = a, b, c, \ldots, k$.

Using the various cumulants and theorems stated earlier, in the results

$$k_1(\psi_i^2) = \sum_i \frac{\sigma_i^2}{m_i} \qquad \ldots(5.4)$$

$$k_2(\psi_i^2) = 2\sum_{i,j} \frac{\sigma^2{}_{ij}}{m_i m_j} \qquad \ldots(5.5)$$

in terms of variances and covariances as given by Patankar (5), we obtain the first two asymptotic moments of $\psi_i^2 (i=1, 2, 3)$ as below :

$$k_1(\psi_1^2) \approx (k-1) + 2\sum_i \sum_{s=1}^{n-1}\left(1-\frac{s}{n}\right)P_{ii}^{(s)} = A_1 \qquad \ldots(5.6)$$

$$k_2(\psi_1^2)\approx 2\Big[(k-1)+4\sum_i\sum_{s=1}^{n-1}\Big(1-\frac{s}{n}\Big)P_{ii}^{(s)}$$

$$+4\sum_i\sum_{s_1,\,s_2=1}^{n-1}\Big(1-\frac{s_1}{n}\Big)\Big(1-\frac{s_2}{n}\Big)P_{ii}^{(s_1+s_2)}\Big]=2B_1 \quad \ldots(5.7)$$

$$k_1(\psi_2^2)\approx(k^2-3)+2\sum_i p_{ii}+2\sum_i\sum_{s=1}^{n-3}\Big(1-\frac{s}{n}\Big)P_{ii}^{(s+1)}=A_2 \quad \ldots(5.8)$$

$$k_2(\psi_2^2)\approx 2[k^2+2k-9+4\sum_i p_{ii}+2\sum_{\substack{i,j\\ i\neq j}} p_{ji}p_{ij}$$

$$+4\sum_i\sum_{s=1}^{n-3}\Big(1-\frac{s}{n}\Big)\Big[p_{ii}^{(s)}+p_{ii}^{(s+1)}+p_{ii}^{(s+2)}\Big]$$

$$+4\sum_i\sum_{s_1,\,s_2=1}\Big(1-\frac{s_1}{n}\Big)\Big(1-\frac{s_2}{n}\Big)P_{ii}^{(s_1+s_2+2)}\Big]=2B_2 \quad \ldots(5.9)$$

$$k_1(\psi_3^2)\approx k^3-5+2\sum_i p_{iii}+2\sum_{i,j} p_{jij}p_{iji}+2\sum_i\sum_{s=1}^{n-5}\Big(1-\frac{s}{n}\Big)P_{ii}^{(s+2)}=A_3 \ldots (5.10)$$

$$k_2(\psi_3^2)\approx 2\Big[k^3+2k^2+2\sum_{m.\,k,\,j}\frac{P_m p_{mi}p_{mij}^2}{p_i p_{ij}}-25+8\sum_i p_{iii}$$

$$+6\sum_{i,j} p_{iji}p_{jij}+4\sum_i p_{iii}^2+4\sum_{\substack{i,j,m\\ i\neq j\neq m}} p_{ijm}\,p_{jmi}\,p_{mij}$$

$$+2\sum_{\substack{i,j,m,r\\ i\neq j\neq m\neq r}} p_{jmr}\,p_{mri}\,p_{rij}\,p_{ijm}+\alpha$$

$$+4\sum_i\sum_{s=1}^{n-5}\Big(1-\frac{s}{n}\Big)\Big\{p_{ii}^{(s+1)}+p_{ii}^{(s+2)}+p_{ii}^{(s+3)}+p_{ii}^{(s+4)}\Big\}$$

$$+4\sum_i\sum_{s_1,\,s_2=1}^{n-5}\Big(1-\frac{s_1}{n}\Big)\Big(1-\frac{s_2}{n}\Big)\Big/P_{ii}^{(s_1+s_2+4)}\Big]=2B_3, \quad \ldots(5.11)$$

where $P_{ii}^{(s+q)}=p_{ii}^{(s+q)}-P_i$ for $q=0, 1, 2, 3, 4$ and $s=s_1+s_2$, $i, j, m, r=a, b, c, \ldots, k$.

$$\alpha=4\sum_i p_{iii}^3+2\sum_i p_{iii}^4+4\sum_{\substack{i,j\\ i\neq j}} p_{iij}\,p_{iji}\,p_{jii}\,(3+2p_{iii})$$

$$+2\sum_{\substack{i,j,m\\ i\neq j,\ m\neq i}} p_{ijm}\,p_{jmi}\,p_{mii}\,P_{iij}$$

$$+2\sum_{\substack{i,j,m\\ i\neq j\neq m}} p_{jim}\,p_{imi}\,p_{mij}\,p_{iji}$$

$$+2\sum_{\substack{i,j,m\\ i,\ m\neq j}} p_{jjm}\,p_{jmi}\,p_{mij}\,p_{ijj}$$

which is negligible for fairly large values of k, since α is the sum of the terms of order

$$\frac{1}{k},\ \frac{1}{k^2} \text{ and } \frac{1}{k^3}.$$

We find that the asymptotic distributions of $\frac{A_i \psi_i^2}{B_i}$ $(i=1, 2, 3)$ can be approximated to the x^2 distribution, as discussed by Patankar[7] with $\frac{A_i^2}{B_i}$ degrees of freedom.

It may be noted that if the sequence is independent the expected values and variances of ψ_1^2, ψ_2^2 and ψ_3^2 reduce to

$$k_1(\psi_1^2)=(k-1)\ ;\ k_2(\psi_1^2)=2(k-1)$$

$$k_1(\psi_2^2)=(k^2-1)\ ;\ k_2(\psi_3^2)=2(k^2+2k-3)$$

$$k_1(\psi_3^2)=(k^3-1)\ ;\ k_2(\psi_2^2)=2(k^3+2k^2+2k-5)$$

$$k_1(\psi_t^2)=(k^t-1)\ ;\ k_2(\psi_t^2)=2(k^t+\frac{2(k^t-1)}{k-1}-2t-1)$$

and they are the same as those derived from the asymptotic distributions by Billingsley[3].

REFERENCES

1. Bartlett, M.S., Proc. Camb. Phil. Soc., Vol. 47(1951), pp. 86-95.
2. Bhat, B.R., *Ann. Math. Stat.* 32, 59 (1961).
3. Billingsley, P., Ann. Math. Stat. 32, 12 (1961).
4. Good, I.J., Ann. Math. Stat., 32, 41(1961).
5. Goodman, Leo A., Ann. Math. Stat., 32, 148(1961).
6. Iyer, P.V.K. and Samarasimhudu, P., Jour, of Ind. Soc. of Agr. Stat., 17 No. 2 (1965).
7. Patankar, V.N., Biometrika, 41, 450 (1954).
8. Whittle. P.J. Roy, Stat. Soc., B.. 17. 235 (1955).

CONTRIBUTIONS TO DESIGN AND ANALYSIS OF EXPERIMENTS—A REVIEW

By

H.K. NANDI*

1. **Introduction.** There is no claim to originality in this article as it is purported to be a review of the work done on some aspects of the design and analysis of experiments, with which the reviewer had an occasion to be associated in some capacity or other. It includes, besides the work of the reviewer, contributions of K.S. Banerjee, K.N. Bhattacharya, P.M. Roy and B. Adhikari. As it is difficult to give a well-connected account of work on diverse topics in the field of design and analysis of experiments, the review will split up into a number of more or less well-defined topics and under each head the work of the authors concerned will be presented. The different heads will be as follows: (*i*) analysis of experiments, (*ii*) balanced incomplete block designs (BIBD), (*iii*) partially balanced incomplete block designs (PBIBD), (*iv*) rectangular lattices, (*v*) weighing designs and (*vi*) balanced block designs with variable replications. It may be noted that in following the above-mentioned order of presentation of topics, no scale of priority or importance is implied, and it is purely a matter of convenience in arranging the topics. The bibliography at the end includes not only the papers of the authors named above but also a few others related to the work under review. Certain standard results and definitions will be assumed throughout this exposition, as their inclusion will make it unnecessarily pedantic.

2. Analysis of experiments

2.1. **General analysis.** In the analysis of any design it was customary to apply an F-test for a group of contrasts and, in case of significance, to apply detailed t-tests for the individual contrasts. The unsatisfactory nature of this procedure has been revealed in a study [37 : (1951)] and a method has been suggested based on the detailed t-statistics right at the start. As is well-known the distribution of the non-central F which determines the power of the test, involves the non-centrality parameter λ, and the two degrees of freedom, k and r one representing the contrasts and the other the error. It has been shown in [37] that the power of the F-test, though increasing with r and λ, is decreasing with k for fixed λ and r. This points out the need of using only a limited number of treatments in any experiment, as otherwise k will increase without increasing λ and thus lower the sensitivity of the experiment. Moreover, as the two-step procedure of first F and then t will affect the conventional error probabilities considerably, a test based on the t's for the independent contrasts was suggested. If $t_i (i=1, 2,...,k)$ are the t-statistics for k independent contrasts, the suggested procedure is as follows :—

1. Reject the null hypothesis when at least one of the t's exceeds c.

---

*Calcutta University.

2. Declare the contrasts corresponding to which t's exceed c as significantly different from 0.

The constant c has to be chosen so as to make the error of first kind equal to a pre-assigned level. It is interesting to note that this method is closely related to the method of simultaneous confidence interval propounded at about the same time by Scheffe' (1953) and Tukey (1951). Other modifications of the F-test have also been considered in [37].

2.2. **Analysis of various types of strip arrangements.** The following model of a multivariate normal population has been studied in [36 ; (1947)]. Let $x_{ijk}(i=1, 2,\ldots,p\ ; j=1. 2,\ldots,q\ ; k=1, 2,\ldots, r)$ follow a pqr-variate normal population with

$$\begin{aligned} E(x_{ijk}) &= m_{ijk} \\ V(x_{ijk}) &= \sigma^2 \\ Cov(x_{ijk}, x_{i'j'k'}) &= \rho_1\sigma^2 \text{ when } i=i', j=j', k\neq k' \\ &= \rho_2\sigma^2 \text{ when } i=i', j=j' \\ &= \rho_3\sigma^2 \text{ when } i\neq i'. \end{aligned}$$

Then by known orthogonal transformations (*i e.*, elements of the transformation matrix are independent of all unknown parameters), they can be reduced to four sets of variables, call them A, B, C and D, containing respectively $pq(r-1), p(q-1), (p-1)$, and 1 variables which are independently normally distributed such that variance remains constant within each set but varies from set to set. Accordingly, tests of linear hypotheses concerning expectations of each set can be performed by the analysis of variance procedure within the set. Best (linear) estimates of linear functions of expectations within the set can also be obtained in the usual way from the variables within the set. These are the so-called intra-set estimate. If the ratios of variances between sets are known the inter-set estimates can be derived in the traditional manner, and finally the pooled estimates. The application of the model to the analysis of incomplete block designs leads to intra-block and inter-block analysis. It also provides justification for the analysis of simple strip or split-plot arrangements and also makes possible the analysis of various complicated types of strip arrangements.

The generation of this model by introducing random effects has also been considered. It may be noted that though only three suffixes i, j, k have been used above, any number of suffixes can be accommodated there in a straightforward manner.

2.3. **Efficiency of experimental designs.** Commonly, the efficiency of a design is related to the average variance of all possible contrasts or a set of independent contrasts as estimated from the random variables of the design. But when the interest of the experimenter lies not in a single contrast at a time but in a number of contrasts simultaneously, the above average loses its relevance. Wald (1943) defined the efficiency of an experimental design by relating it to the value of the determinant of the variance–covariance matrix of the estimates in which one is simultaneously interested, and proved the efficient character of the Latin square and hyper-Graeco square designs. In [38 ; (1951)] a justification is provided for Wald's definition of efficiency from the consideration of power of the relevant F-test. A further justification has been given in [39, (1961)] in terms of the properties of the simultaneous confidence intervals. It is shown that the efficiency of the design so defined attains its maximum when and only when it is an orthogonal one. So Wald's results follow. In the symmetrical balanced incomplete block designs, it follows that the most efficient way of eliminating the effects due to the positions of the plots within a block is the Youden square arrangement.

3. Balanced Incomplete Block Designs (BIBD)

3.1. **Non-isomorphic solutions.** Various methods of construction of BIBD's are well-known. The direct method of construction of such designs from definition, though laborious, yields all possible non-isomorphic solutions and also establishes the non-existence of any such solution. Fisher (1940) undertook the examination of all possible solutions of a particular design : $v=15$, $b=35$, $r=7$, $k=3$, $\lambda=1$. Q.M. Hussain (1945) and (1946) systematized the method of construction of all non-isomorphic solutions of symmetrical BIBD's with $\lambda=2$ by introducing what he called λ-chains. Taking $k=5$ and starting with an initial block containing the treatments 1, 2, 3, 4 and 5, the remaining ten blocks must contain each of the 10 possible pairs of these 5 treatments. Denote these blocks by the pairs of treatments they contain. Then the placement of a new treatment 6 in these ten blocks can be written in the form of a chain like (12345), meaning thereby that the blocks containing the treatment 6 are (12), (23), (34), (45), (51). If $k=6$, it is found that only two types of chains are relevant *viz.* (*abcdef*) or (*abc*)(*def*). Enumerating all possible consistent chains he proved that (*a*) $v=b=11$, $r=k=5$, $\lambda=2$ has a unique solution except for permutations of the treatments which render the solutions isomorphic to one another ; (*b*) $v=b=16$, $r=k=6$, $\lambda=2$ has three non-isomorphic solutions, and (*c*) $v=b=22$, $r=k=7$, $\lambda=2$ has no solution.

If the BIBD is not symmetrical, the enumeration of non-isomorphic solutions is somewhat more difficult, as the blocks of the design may have different numbers of common treatments between themselves, instead of a constant number as in the case of symmetrical BIBD. If the frequency distribution of common treatments between an initial block and any one of the rest is known, the enumeration work can be undertaken on some systematic basis. In [34 ; (1946)] this frequency distribution has been derived as solutions of a set of linear equations. This distribution is unique for $\lambda=1$ and 2. It is found that the asymmetrical designs ; A (6, 10, 5, 3, 2) and B (10, 15, 6, 4, 2) have respectively one and three non-isomorphic solutions, which are isomorphic to those obtained by "residuation" from the corresponding symmetrical designs considered by Hussain.

This raised the following interesting problem. Let (v, v, r, r, λ) be a symmetrical BIBD. From this can be obtained the residual design : $(v-k, v-1, r, r-\lambda, \lambda)$

by the process of residuation, *i.e.* by suppressing one block and all the k treatments occurring in that block from the remaining $v-1$ blocks, and the derived design, $(k, v-1, r-1, \lambda, \lambda-1)$ by the process of derivation, *i.e.* by supressing one block and all treatments not occurring in that block from the remaining $v-1$ blocks. Given any symmetrical BIBD, residuation and derivation are always possible and they yield unique BIBD. What about the inverse operation, say, integration? Is it always possible to integrate a residual and a corresponding derived into a symmetrical BIBD? It is easily shown by considering the residual design (16, 24, 9, 6, 3) given by Bhattacharya [14 ; (1944)] and any derived design (9, 24, 8, 3, 2) that integration is not possible to yield a solution for the symmetrical design (25, 25, 9, 9, 3). In [35 ; (1946)] the non-isomorphic solutions of the residual (8, 14, 7, 4, 3) and the derived (7, 14, 6, 3, 2) are obtained and they are integrated to yield solutions of the symmetrical design (15, 15, 7, 7, 3). It is found that the residual and the derived have four non-isomorphic solutions each whereas the symmetrical have five non-isomorphic solutions, thus demonstrating that all possibilities are open for integration. From the unique frequency distribution of common treatments in the case of $\lambda=1$ (referred to above) it can be easily shown that integration is uniquely possible for $\lambda=1$. For the case $\lambda=2$, the possibility of integration of (15, 21, 7, 5, 2) and (7, 21, 6, 2, 1) is shown in general in [33 ; (1945)] and thus the impossibility of (15, 21, 7, 5, 2) is proved as the symmetrical (22, 22, 7, 7, 2) is known to be impossible. For $\lambda=2$ it was conjectured that integration was always possible, and this conjecture was proved to be true by Hall and Connor (1954). This result taken along with the result of [33] shows that this integration is unique (but for permutation of the treatments).

3.2. **New solutions of BIBD's.** Bhattacharya derived by following ingenious methods the first solution of the following designs ; (29, 29, 9, 9, 3) and (31, 31, 10, 10, 3) and their residuals (20, 28, 9, 6, 3) and (21, 30, 10, 7, 3). [13 ; (1944) ; 14 ; (1944) ; 15 ; (1946)].

4 Partially Balanced Incomplete Block Designs (PBIBD)

4.1. **Relation between PBIBD and other incomplete block arrangements.** Roy's study in [43, (1954) ; 45, (1954) ; 45, (1954) ; 47, (1955)] reveals interesting correspondence between PBIBD and BIBD, and PBIBD and other incomplete block arrangements. It is shown that BIBD's of the following types :— (*i*) with $\lambda=1$ and $v \neq b$, (*ii*) obtained by residuation (or block-section) from symmetrical BIBD's, (*iii*) affine resolvable with $\lambda=nt+1$ and (*iv*) obtained by taking all groups of k varieties out of v varieties—lead to PBIBD's on dualisation. Further, the complementary BIBD's to the above give PBIBD's on dualisation. It may be noted that some of these results were obtained independently by Shrikhande (1952) at about the same time. Roy defines an affine resolvable arrangement without requiring the condition of balancing as an arrangement of v elements in b blocks of k elements each such that each element occurs in r blocks and the b blocks can be separated into r sets of n blocks each, each set of n blocks containing all the v elements just once (*i.e.* forming a complete replication of the v elements) and each block of a set having an equal number of elements in common with any other block not in the set. The dualisation of this affine resolvable arrangement or that of a sub-set of complete replications of this arrangement always results in PBIBD. He has also investigated the conditions under which a part of a BIBD or a combination of a BIBD and PBIBD will give PBIBD. As an illustration, consider a BIBD with $v=n^2m$, $b=nr$, $k=nm$, λ exists. Then there is a 3-associate PBIBD with the following parameters :

$$v^*=n^2m, \quad b^*=nr+2nt, \quad r^*=r+2t, \quad k^*=nm$$

$$n_1=m-1, \quad n_2=2m(n-1), \quad n_3=m(n-1)^2$$

$$\lambda_1=\lambda+2t, \qquad \lambda_2=\lambda+t \qquad \lambda_3=\lambda,$$

$$(p^1{}_{jk})=\begin{bmatrix} m-2 & 0 & 0 \\ 0 & 2m(n-1) & 0 \\ 0 & 0 & m(n-1)^2 \end{bmatrix}$$

$$(p^2{}_{jk})=\begin{bmatrix} 0 & m-1 & 0 \\ m-1 & m(n-2) & m(n-1) \\ 0 & m(n-1) & m(n-1)(n-2) \end{bmatrix}$$

$$(p^3{}_{ik})=\begin{bmatrix} 0 & 0 & m-1 \\ 0 & 2m & 2m(n-2) \\ m-1 & 2m(n-2) & m(n-2)^2 \end{bmatrix}$$

This PBIBD can be obtained by adding to the BIBD t times the two sets of n blocks formed in the following manner. Distribute then n^2m varieties of the design in the n^2 cells of an $n \times n$ square so that every cell gets m varieties. Taking the rows and columns of this square as blocks one gets the two sets of n blocks which themselves form a PBIBD.

4.2. **Generalization of group divisible designs (GDD).** Roy [42, 1953); 47, (1955); 47, (1955); 50,(1962)] explored the possibility of generlizing the concept of two-associate group divisible designs of Bose and Connor (1952) in various directions to obtain new sub-classes of PBIBD's. In two-associate GDD, the varieties are divided into groups, each group containing the same number of varieties, and any two varieties belonging to two different groups are called one type of associate—call them first associates while any two varieties within a group are second associates. Now each group may be further divided into sub-groups and between sub-group variety pairs may be called second associates and within sub-group varieties pairs third associates. and so on. Thus is obtained the *Hierarchical Group Divisible designs* with m associate classes (HGD_m)[42, 47, 50]. They may be precisely defined in the following manner.

An incomplete block design with v varieties arranged in b blocks of size k each in such a way that each variety is replicated r times occurring at most once in a block, is called HGD$_m$ if the v varieties can be divided into $N=N_1.N_2..\ N_{m-1}$ hierarchical groups of N_m varieties each as follows :

At the first stage the v varieties can be divided into N_1 groups of $S_1=N_2.N_3$, .. . N_m varieties each such that any two varieties not belonging to the same group are first associates and occur together in just λ_1 blocks. At the second stage S_1 varieties of any group of the first stage can be further subdivided into N_2 groups of $S_2=N_3$, N_4 . , N_m varieties each such that any two varieties of a group of the first stage but not belonging to the same group of the second stage are second associates and occur together in just λ_2 blocks. Proceeding in this way at the $(m-1)$ th stage $S_{m-2}=N_{m-1}$ N_m varieties of any group of the $(m-2)$ th stage can be further subdivided into N_{m-1} groups of N_m varieties each such that any two varieties of a group of the $(m-2)$-th stage but not belonging to the same group of the $(m-1)$-th stage are $(m-1)$ th associates and occur together in just λ_{m-1} blocks. Lastly, any two varieties belonging to the same group of the $(m-1)$-th stage are m-th associates and occur together in just λ_m blocks. It has been shown that this class of designs (HGD$_m$) forms a sub-class of the PBIB designs with the following parameters, and conversely the PBIB designs with these parameters uniquely determine the association scheme.

$$v=N_1N_2...N_m,\ b,\ r,\ k,\ \lambda_i,$$

$$n_i=N_iN_{i+1}...N_m-N_{i+1}\ N_{i+2}...N_m\ ;\ i=1,\ 2...,\ m$$

$$(p^{t}_{jk})=\begin{array}{r} \\ \\ \\ i\text{-th row}\rightarrow \\ \\ \\ \\ \end{array}\left[\begin{array}{ccccccc} S_1(N_1-1) & 0 & \ldots\ldots & 0 & 0 & \ldots\ldots & 0 \\ 0 & S_2(N_2-1) & \ldots\ldots & 0 & 0 & \ldots\ldots & 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & 0 & \ldots S_i(N_i-2) & S_{i+1}(N_{i+1}-1) & \ldots & & S_m(N_m-1) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & 0 & S_{i+1}(N_{i+1}-1) & 0 & \ldots\ldots & & 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & 0 & S_m(N_{m-1}) & 0 & \ldots\ldots & & 0 \end{array}\right]$$

where $S_l=N_{l+1}\ N_{l+2}\cdots N_m,\ l=1,\ 2,...m-1$ and $S_m=1$.

A second generalization of the GD association scheme would be to consider simultaneously alternate groupings of the varieties [46, (1954), 47, (1955)]. Let $v=pq$ varieties be arranged in the form of a $p\times q$ rectangle and consider the two alternate groupings—one by rows and the other by columns of the rectangle. A variety pair is a first associate if they belong to the same rows, a second associate if they belong to the same column and a third associate if they belong neither to the same row nor to the same column. If $v=p^2$, we have a square on which can be imposed orthogonal

Latin squares and alternate groupings are possible by rows, columns and the letters of any number of orthogonal Latin squares. Due to the orthogonal nature of the alternate groupings, the association scheme has been called as orthogonal group divisible association scheme. More precisely, an *Orthogonal Group Divisible* design with $i+1$ associates (OGDL$_i$) is defined as an incomplete block design with $v=n^2$ varieties each replicated r times in b blocks of size k such that the n^2 varieties can be divided into i orthogonal sets of n groups of size n each and any variety pair belonging to the same group of the j-th set, called j-th associate, occurs together in λ_j blocks ($j=1, 2,..., i$) and a variety pair not belonging to the same group of any of the i orthogonal sets, to be called ($i+1$)-th associate, occurs together in just λ_{i+1} blocks. It has been proved that these OGDL$_i$'s form a sub-class of PBIB designs with the following parameters and, conversely, PBIB's with these parameters uniquely, determine the association scheme.

$$v=n^2,\ b,\ r,\ k,\ \lambda_1,\ \lambda_2,\ldots\lambda_{i+1},$$

$$n_j=n-1,\ j=1,\ 2,\ldots,i,\ n_i+1=(n-1)(n-i+1).$$

$$\left(p^{1}_{jk}\right)=\begin{bmatrix} n-2 & 0 & 0\ldots0\ldots & 0 & 0 \\ 0 & 0 & 1\ldots1\ldots & 1 & n-i+1 \\ 0 & 1 & 0\ldots1\ldots & 1 & n-i+1 \\ \hdotsfor{5} \\ 0 & 1 & 1..\ 0\ldots & 1 & n-i+1 \\ \hdotsfor{5} \\ 0 & 1 & 1\ldots1\ldots & 0 & n-i+1 \\ 0 & n-i+1 & n-i+1..\ n-i+1\ldots & n-i+1 & (n-i)(n-i+1) \end{bmatrix}$$

$$\left(p^{l}_{jk}\right)=\begin{bmatrix} 0 & 1 & 1\ldots1 & 0 & 1\ldots & 1 & n-i+1 \\ 1 & 0 & 1\ldots1 & 0 & 1\ldots & 1 & n-i+1 \\ 1 & 1 & 0\ldots1 & 0 & 1\ldots & 1 & n-i+1 \\ \hdotsfor{7} \\ 1 & 1 & 1\ldots0 & 0 & 1\ldots & 1 & n-i+1 \\ 0 & 0 & 0\ldots0 & n-2 & 0\ldots & 0 & 0 \\ 1 & 1 & 1\ldots1 & 0 & 0\ldots & 1 & n-i+1 \\ \hdotsfor{7} \\ 1 & 1 & 1\ldots1 & 0 & 1\ldots & 0 & n-i+1 \\ n-i+1 & n-i+1 & n-i+1\ldots n-i+1 & 0 & n+i+1\ldots & n+i+1 & (n-i)(n-i+1) \\ l=2,\ 3,\ldots i, & & & & & & \end{bmatrix}$$

$$\left(p_{jk}^{i+1}\right)=\left[\begin{array}{cccccccc} 0 & 1 & 1 & \ldots & 1 & \ldots & 1 & n-i \\ 1 & 0 & 1 & \ldots & 1 & \ldots & 1 & n-i \\ 1 & 1 & 0 & \ldots & 1 & \ldots & 1 & n--i \\ \multicolumn{8}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ 1 & 1 & 1 & \ldots & 0 & \ldots & 1 & n-i \\ \multicolumn{8}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ 1 & 1 & 1 & \ldots & 1 & \ldots & 0 & n-i \\ n-i & n-i & n-i & \ldots & n-i & \ldots & n-i & (n-i)^2+i-2 \end{array}\right]$$

In the papers and the thesis referred to above Roy has also studied the properties of these designs and given methods of construction of some series of designs belonging to these classes.

4.3. Generlization of two-associate cyclical association scheme.

Adhikari [3, (1965)] has considered some generalization of the two associate cyclical association scheme of Bose and Shimamoto (1952). The two-associate cyclical asso- ciation scheme of Bose and Shimamoto can be regarded as an abelian group G of v elements such that the non-unit elements of G can be divided into disjoint subsets A and B *i.e.* $G-1=AUB$ where the n elements of A are such that among the $n(n--1)$ non-unit ratios arising out of them, the elements of A are each repeated α times and those of B β times. Then the treatment corresponding to the element θ has its first associates as θA and second associates as θB. Generalizing this idea, consider an abelian group G consisting of n elements. Let it be possible to decompose G into direct factors : $G=G_1\times G_2\times\ldots\times G_p$, G_i containing m_i elements. Let among this set of p groups, a subset of l groups, say i_1-th...i_l-th groups, admit the following further decomposition $G_{ij}-1=A_{ij}\,UB_{ij}, j=1, 2,\ldots,l$, where A_{ij} consists of m'_{ij} elements and B_{ij}. m''_{ij} elements. Let, further, the elements of A_{ij} be such that among the non-unit ratios arising out of them, the elements of A_{ij} appear α_{ij} times each and the elements of B_{ij}, β_{ij} times each. Obviously, we must have $m'_{ij}\alpha_{ij}+m''_{ij}\ \beta_{ij}=m'_{ij}(m'_{ij}-1)$. Without loss of generality we assume $i_1<i_2\ldots<i_l$.

Let the first associates of the treatment corresponding to the element θ be $\theta(G_1-1)$, its second associates $\theta G_1\times(G_2-1)$...its (i_1-1)-th associates $\theta G_1\times G_2\times\ldots\times(G_{i_1-1}-1)$, its i_1-th associates $\theta G_1\times G_2\times\ldots\times G_{i_1-1}\times A_{i_1}$, its (i_1+1)-th associates $\theta G_1\times G_2\times\ldots\times G_{i_1}-1\times B_{i_1}$, its (i_1+2)-th associates $\theta G_1\times G_2\times\ldots\times G_{i_1}\times(G_{i_1+1}-1)$, and so on. The parameters of the association scheme will be easily seen to be as follows.

$$n_1=m_1-1,\ldots,n_{i_1}-1,\ m_1m_2\ldots m_{i_1}-2(m_{i_1-1}-1),$$

$$n_{i_1}=m_1m_2\ldots m_{i_1-1}\,m'_{i_1},\ n_{i_1+1}=m_1m_2\ldots m_{i_1-1}\,m''_{i_1}.$$

$$n_{i_1+2}=m_1m_2\ldots m_{i_1-1}\,m_{i_1}\,(m_{i_1+1}-1),\ldots$$

$$n_{l+p}=m_1m_2\ldots m_{p-1}(m_p-1).$$

$$\left(p_{jk}^{w}\right)=\begin{array}{r} \\ \\ \\ \\ w\text{-th row} \\ \\ \\ \\ \end{array}\begin{bmatrix} 0 & 0 & \cdots & 0 & n_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & n_2 & 0 & \cdots & 0 \\ \hdotsfor{8} \\ 0 & 0 & \cdots & 0 & n_{w-1} & 0 & \cdots & 0 \\ n_1 & n_2 & \cdots & n_{w-1} & m_1\ldots m_{w-1}(m_{w-2}) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & n_{w+1} & \cdots & 0 \\ \hdotsfor{8} \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & n_{l+p} \end{bmatrix}$$

for $w=1, 2, \ldots, l+p$ except $i_1, i_1+1, i_2+1, i_2+2, \ldots i_l+(l-1), i_l+l$.

$$\left(p_{jk}^{t}\right)=\begin{bmatrix} 0 & 0 & \cdots & 0 & n_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & n_2 & 0 & 0 & \cdots & 0 \\ \hdotsfor{9} \\ n_1 & n_2 & & n_{t-1} & m_1m_2\ldots m_{t-1}\alpha_t & \begin{matrix} m_1m_2\ldots m_{t-1} \\ (m_t-\alpha_t-1) \end{matrix} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \begin{matrix} m_1m_2\ldots m_{t-1} \\ (m_t-\alpha_t-1) \end{matrix} & \begin{matrix} m_1m_2\ldots m_{t-1} \\ (m_{t+1}-m_t+\alpha_t+1) \end{matrix} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & n_{t+2} & \cdots & 0 \\ \hdotsfor{9} \\ 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & n_{1+p} \end{bmatrix}$$

for $[t, \alpha_t)=(i_1, \alpha_{i_1}), (i_2+1), (i_2+1, \alpha_{i_2}), \ldots, (i_1+l-1, \alpha_{i_l})$

$$\left(p_{jk}^{t+1}\right)=\begin{array}{r} \\ \\ \\ \\ \\ (+1)\text{-th row} \\ \\ \\ \\ \end{array}\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & n_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & n_2 & 0 & \cdots & 0 \\ \hdotsfor{9} \\ 0 & 0 & \cdots & 0 & 0 & n_{t-1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & m_1m_2\ldots m_{t-1}\beta_t & \begin{matrix} m_1m_2\ldots m_{t-1} \\ (m_t-\beta_t) \end{matrix} & 0 & \cdots & 0 \\ n_1 & n_2 & \cdots & n_{t-1} & \begin{matrix} m_1m_2\ldots m_{t-1} \\ (m_t-\beta_t) \end{matrix} & \begin{matrix} m_1m_2\ldots m_{t-1} \\ (m_{t+1}-m_t-\beta_t-1) \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \cdots & \begin{matrix} 0 \\ 0 \end{matrix} \\ 0 & 0 & \cdots & 0 & 0 & 0 & n_{t+2} & \cdots & 0 \\ \hdotsfor{9} \\ 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & n_{l+p} \end{bmatrix}$$

for $(t, \beta_t)=(i_1, \beta_{t1}), (i_2+1, \beta_{i2}), \ldots\ldots, (i_1+l-1, \beta_{tl})$

The properties of this sub-class of PBIBD's have also been studied in [3].

5. **Rectangular Lattices.** Lattice designs of Yates (1936, 1937) are well-known. These however require the number of varieties to be an exact power of the block size. Obviating this restriction, Harshbarger (1947, 1949, 1951) introduced rectangular lattices for varieties of the form $p(p-1)$ by writing the varieties in a $p \times p$ square from which the diagonal cells are missing and choosing blocks corresponding to rows, columns or letters of a Latin square superposed on the $p \times p$ square. Accordingly, he had simple lattice and triple lattice, and he further defined somewhat vaguely near balance lattice Nair (1951, 1953) has discussed these designs from the view point of PBIBD, showing that simple and near balance rectangular lattices belong to the system of PBIBD's with four ($p \geqslant 4$) and two associates classes respectively, whereas triple rectangular lattice does not belong to this system. He has further observed that $p \times (p-1)$, $(p-1)$-ple rectangular lattice is a PBIBD with three associate classes and the dual of a $p \times (p-1)$, p-ple rectangular lattice is a square lattice with $(p-1)$ replications whereas the dual of any other rectangular lattice is a PBIBD with three associate classes.

Roy [46, (1954) ; 47, (1955) ; 48, (1955) ; 49, (1957)] has made precise the definitions of rectangular lattices of various types of such designs. The rectangular lattices possess the dual properties mentioned by Nair (1953) in general only when they are latinized. $p \times (p-l)$, n-ple latinized rectangular lattice is defined to be an arrangement of $p \times (p-l)$ varieties in n resolvable replicates of blocks of size $(p-l)$ such that any pair of varieties occurs not more than once in the whole design and that by adding lp new varieties a n-ple square lattice can be formed. $p \times (p-l)$, p-ple as well as $(p-1)$-ple latinized rectangular lattices with block size $(p-l)$ for $l > 1$ are available and are capable of easy analysis. A new method for identification of different associates has been given with the help of OGD designs discussed earlier. For $n > p/2$, a most convenient alternative method of determining the associates has also been found. A simplified analysis of such designs has been worked out providing the expressions of two common variances—one for comparing two varieties occurring together in some block and the other for two varieties not occurring together in any block. These common variances are properly weighted averages of variances of different varietal differences, based on the distribution of varieties among the different associate classes.

6. **Weighing Designs.** Hotelling (1944) posed the problem of weighing a number of objects in suitable combinations to attain maximum accuracy of measurement with a given number of weighing operations. Suppose there are p objects whose true weights $\beta_1, \beta_2, \ldots\ldots; \beta_p$ are to be estimated by N weighing operations. Associate with the i-th object a variable $x_{\lambda i}$ which takes the value $+1$, -1, or 0 according as the object is placed on the left, right or neither pan of the balance in the λth weighing operation. If $x_{\lambda i}$'s are restricted to values 1 and 0, the corresponding weighing problem is known as the spring balance problem as distinguished from the earlier problem usually referred to as the chemical balance problem. The $N \times p$ matrix $X = (x_{\lambda i})$ is known as the design matrix. From the least square theory it follows that the dispersion matrix of the best linear estimates, $A = (a^{ij})$ is $(X'X)^{-1}$. a^{ii} are called the variance factors of the estimates.

Hotelling (1944) showed that the best design matrix will be one for which A is a diagonal matrix with each diagonal element equal $1/N$. Kishen (1945) constructed such optimum chemical balance designs for $N=2^m$ (m is a positive integer) with the help of hyper-graeco latin cubes. Mood (1946) showed that when $N=0$ (mod 4), the optimum chemical balance design can always be derived from a Hadamard matrix H_N. Plackett and Burman (1943-46) have constructed all Hadamard matrices upto the order 100 (except 92).

If a best design in Hotelling's sense does not exist, Mood (1946) defined an efficient design as one for which the determinant A has the least value. For the chemical balance problem when a Hadamard matrix does not exist, Mood showed that for $N=i$ (mod 4), $i=1, 2, 3$, quite efficient designs can be constructed from H_{4k} by addition of suitable rows. In [7, (1949)] Banerjee studied the effect of adding different rows to H_{4k} by evaluating the variance factors. For the spring balance problem, Mood (1946) has derived efficient designs (L_N) from H_{N+1} when $N=p$, and suggested a combination of P_k matrices (*i e.* $p_k \times p$ matrices whose rows are all the arrangements of k ones and $p-k$ zeros) when $N>p$. In [6, (1948) ; 8, 1949) ; 9, (1949)], Banerjee has shown that balanced incomplete block designs (BIBD) also furnish efficient weighing designs in Mood's sense and symmetrical BIBD's correspond to the class of L_N designs. The BIBD has the advantage of a constant variance factor and a constant covariance between any pair of estimates. The variance factors of designs obtained from addition or deletion of rows from L_N have been calculated in [9, (1949)] and comparative study of the designs has been made. He has further shown [10, (1950)] that by modifying the BIBD it is possible to have an efficient weighing design in many cases of biased spring balance, which supplies orthogonal estimates. In [12, (1951)] he has shown how partially balanced incomplete block designs (PBIBD's) may be used as efficient spring balance designs. Developing a result of Kempthorne (1948) on the construction of efficient chemical balance designs with the help of fractional replicates of 2^n factorial designs, Banerjee [8, (1949)] has proved that when a 3/4 replicate is used as a chemical balance design, the variance factors come out as $1/2^{n-1}$, and that in a weighing design given by fractional replicate of the type $(2^\beta - 1)/2^\beta$ $(1 \leqslant \beta \leqslant n)$ of a 2^n experiment, the variance factor for each of the estimates is $1/2^{n-1}$. In [9, (1949)] he has considered a 3/4 replicate as a spring balance design and has found that the variance factor for all the objects except those denoted by the defining set is $1/2^{n-3}$ which is 1/4 that in the chemical balance problem. He has also drawn attention to the fact that an optimum design for estimating the individual weights is not necessarily optimum for estimating a linear function of weights. Efficient designs for the latter type of problems can also be obtained from BIBD's [6, (1948)].

7. Balanced Block Designs with Variable Replications.

Since Nair and Rao (1942) proposed the inter and intragroup balanced block designs, there has been no attempt for the construction of these designs and the study of its properties. Corsten (1962) first considered the problem of constructing a balanced block design with two different replications. Agarwal and Raghavachari

[4, (1963) ; 5, (1964] have considered construction of such designs with three different replications. In [1, (1965)] Adhikari studied the properties of such designs and discussed various methods of their construction.

Adhikari derived a number of useful inequality relations among the parameters of those designs and, with their help, proved the non-existence of many such designs. For designs with two replications he explored the possibility of attaining equal error variance for both the intra-group contrasts different from that of the inter-group contrasts, or, alternatively, equal error variances for one intra-group and the inter-group contrasts different from that of the other intra-group contrasts. He has enumerated the parameters of all such designs with r_1, $r_2<20$, $\lambda_{ij}<10$, $k<10$, and have shown some of them to be impossible and have constructed the others. He has also described various methods of construction of balanced block designs with variable replications in general. In [2, (1965)] he has given a useful difference theorem for such construction.

REFERENCES

1. Adhikari, Basudev (1965). On the properties and construction of balanced block designs with variable replications. Cal. Stat. Assoc. Bull, 14, 36-64.
2. Adhikari, Basudev (1965). A difference theorem for the construction of balanced block designs with variable replications. *Ibid*, 14,... .
3. Adhikari, Basudev (1965). Generlization of 2-associate cyclical association scheme to higher associate classes. Communicated to the Indian Science Congress 53rd Session.
4. Agrawal, H. (1963). On balanced block designs with two different numbers of replications. Jour. Ind. Stat. Assoc., 1, 145-151.
5. Agrawal, H. and Raghavachari, R. (1964). On balanced block designs with three different numbers of replications. Cal. Stat. Assoc. Bull.. 13, 80-86.
6 Banerjee, K.S. (1948). Weighing designs and balanced incomplete blocks. Ann. Math. Stat., 19, 394-399.
7. Banerjee, K.S. (1949). On the variance factors of weighing designs in between two Hadmard matrices. Cal. Stat. Assoc. Bull., 2, 38-42.
8. Banerjee, K.S. (1949). A note on weighing designs. Ann. Math. Stat., 20, 300-304.
9. Banerjee, K.S.(1949). On certain aspects of spring balance designs· Sankhya, 9, 367-376.
10. Banerjee, K.S. (1950). How balanced incomplete block designs may be made to furnish orthogonal estimates in weighing designs. Biometrika, 37, 50-58.
11. Banerjee, K. S.(1950). Weighing designs. Cal. Stat. Assoc. Bull., 3, 64-76.
12. Banerjee, K. S. (1951). Weighing designs and partially balanced incomplete blocks. *Ibid*, 4, 36-38.
13. Bhattacharya, K.N. (1944). On a symmetrical balanced incomplete block design. Bull. Cal. Math. Soc., 36, 91-96.
14. Bhattacharya, K.N.(1944). A new balanced incomplete block design. Science and Culture, 9, 508.

15. Bhattacharya, K.N. (1946). A new solution in symmetrical balanced incomplete block designs ($v=b=31, r=k=10, \lambda=3$). Sankbya, 7, 423-424.

16. Boss. R.C. and Connor, W.S. (1952). Combinatorial properties of group divisible incomplete block designs. Ann. Math. Stat., 23, 367-383.

17. Bose, R.C. and Shimamoto, T. (1952). Classification and analysis of partially balanced incomplete block designs with two associate classes. Jour. Am. Stat. Assoc., 47, 151-184.

18. Corsten, L.C.A. (1962). Balanced block designs witb two different numbers of replications. Biometrics, 18, 499-519.

19. Fisber, R.A. (1940). An examination of the different possible solutions of a problem in incomplete blocks. Ann. Eug., 10, 52-75.

20. Hall, Marshall (Jr.) and Connor, W.S. (1954). An embedding theorem for balanced incomplete block design. Cand. Jour. Math. 6, 35-51.

21. Harshbarger, B. (1947). Rectangular lattices. Virginia Agricultural Experiment Station, Memoir 1.

22. Harshbarger, B. (1949). Triple rectangular lattices. Biometrics, 5, 1-13.

23. Harshbarger, B. (1951). Near balance rectangular lattices. The Virginia Jour. Sc., 2 (New series), 13-27.

24. Hotelling, H. (1944). Some improvements in weighing and other experimental techniques. Ann. Math. Stat. 15, 297-306.

25. Hussain, Q.M. (1945). On the totality of the solutions for the symmetrical incomplete block designs. $\lambda=2$, $k=5$, or 6. Sankhya, 7, 204-208.

26. Hussain Q.M. (1946). Impossibility of the symmetrical incomplete block design with $\lambda=2$, $k=7$. Sankhya, 7, 317-322.

27. Kamptborne, O. (1948) The factorial approach to the weighing problem. Ann. Math. Stat 19, 238-245.

28. Kishen, K. (1945) On the design of experiments for weighing and making other types of measurement. Ann. Math. Stat., 16, 294-300.

29. Mood, A.M. (1946). On Hottelling's weighing problem. Ann. Matb. Stat., 17, 432-446.

30. Nair, K.R. (1951). Rectangular lattices and partially balanced incomplete block designs. Biometrics, 7. 145-154.

31. Nair, K.R. (1953). A note on rectangular lattices. Biometrics, 9, 101-106.

32. Nair. K.R. and Rao, C.R. (1942). Incomplete block designs for experiments involving several groups of varieties. Science and Culture, 7, 615-616.

33. Nandi, H.K. (1945). On the relation between certain types of tactical configurations. Bull. Cal. Math. Society, 37, 92-94.

34. Nandi, H.K. (1946). Enumeration of non-isomorphic solutions of balanced incomplete block designs. Sankhya, 7, 305-312.

35. Nandi, H.K. (1946). A further note on non-isomorphic solutions of incomplete block designs. Ibid. 7, 313-316.

36. Nandi, (H.K. 1947). A mathematical set-up leading to analysis of a class of designs. Sankhya, B, 172-176,

37, Nandi, H.K. (1951). On analysis of variance test, Cal. Stat. Assoc. Bull.. 3, 103-114.

A STUDY THROUGH A MARKOV CHAIN OF A POPULATION UNDERGOING CERTAIN MATING SYSTEMS IN THE PRESENCE OF LINKAGE

By

PREM S. PURI*

1. **Introduction.** How the distribution of various genotypes and its properties vary from one generation to another in a population undergoing a given system of mating has been and perhaps still is the subject of extensive study among several research workers. A great deal concerning this can be found in literature (see Kempthorne [4]). There are left however a great number of cases which are either still unsolved or are such that their properties have not been fully explored. One such case forms the subject of this paper. One of the key assumptions made in most of the cases studies thus far is that of independent segregation of the factors involved. The only case known in literature where the two factors are assumed to be linked, is the one where the population is subjected to random mating. In the present paper an attempt is made to eliminate the assumption of independent segregation among factors.

To begin with, we study the case of two linked factors for a diploid population undergoing selfing starting with an arbitrary initial genotypic distribution. Later this is generalised to a case of great interest to plant breeders namely that of mixed self-fertilization and random mating. Here it is assumed that the initial population is in equilibrium with respect to random mating system. The case with an arbitrary initial genotypic distribution is somewhat involved and will be reported elsewhere in order not to overload a single paper (see Puri [7]). The populations undergoing such a mixed mating system have been studied by several research workers in the past and more recently by Ghai ([1].[2]), all under the assumption of independent segregation of the factors involved

The model considered here is still restrictive in the sense that it assumes an absence of selection and mutation and assumes also an equal mortality and fertility over all genotypes. The model incorporating selection presents interesting problems and is left for a later study. Reader is referred to an interesting recent paper on selection by *Li* [5].

2.0. **A Markov Chain related to model where population is undergoing continued selfing.** Consider the case of a diploid population with two linked factors each with two alleles and let $A-a$ and $B-b$ be the corresponding gene-pairs with $p(q=1-p)$ denoting the usual recombination value or the proportion of cross-over gametes. Following the standard convention we restrict p to take values between 0 and $\frac{1}{2}$. The case with $p=0$ is that of complete linkage and p is $\frac{1}{2}$ when the factors segregate independently. The four possible gametes are AB, Ab, aB, and ab and the ten possible genotypes are

$$(1)\quad \left\{\begin{array}{lllll} 1.\ AB/AB & 2.\ AB/Ab & 3.\ Ab/Ab & 4.\ AB/aB & 5.\ AB/ab \\ 6.\ Ab/aB & 7.\ Ab/ab & 8.\ aB/aB & 9.\ aB/ab & 10.\ ab/ab. \end{array}\right.$$

Purdue University, Lafayette, Indiana.

38, Nandi, H.K. (1951). On the efficiency of experimental designs. Ibid, 3, 167-171.

39. Nandi, H.K. (1961). Some aspects of simultaneous confidence interval estimation. Ibid, 10, 131-138.

40. Plackett, R.L. and Burman. J.P. (1943-46). The design of optimum multi-factorial experiments. Biometrika, 33, 305-325.

41. Roy, P.M. (1952). A note on the resolvability of balanced incomplete block designs. Cal. Stat. Assoc. Bull., 4, 130-132.

42. Roy, P.M. (1953). Hierarchical group divisible incomplete block designs with *m* associate classes. Science and Culture, 19, 210-211.

43. Roy, P.M. (1954). On the method of inversion in the construction of partially balanced incomplete block designs from the corresponding BIB designs. Sankhya, 14, 39-52.

44. Roy, P.M. (1954). Inversion of incomplete block designs. Bull. Cal. Math. Soc., 46, 47-58.

45. Roy, P.M. (1954). On the relation between BIB and PBIB designs. Jour. Ind. Soc. Agr. Stat., 6, 30-47.

46. Roy, P.M. (1954). Rectangular lattices and orthogonal group divisible designs. Cal. Stat. Assoc. Bull., 5, 87-98.

47. Roy, P.M. (1955.) On some combinatorial problems in the design of experiments. D. Phil. thesis (unpublished). Calcutta University.

48. Roy, P.M. (1955). Analysis of $p \times (p-1)$ *n*-ple latinized rectangular lattices and their multiples. Cal. Stat. Assoc. Bull., 6, 113-131.

49. Roy, P.M. (1957). On the distribution of varieties of $p(p-1)$ *n*.ple latinized rectangular lattices and weighted average variances. Ibid, 7, 101-114.

50. Roy, P.M. (1962). On the properties and construction of HGD designs with *m* associates classes. Ibid, 11, 10-38.

51. Scheffe', H. (1953). A method for judging all contrasts in the analysis of variance. Biometrika, 40, 87-104.

52. Shrikhande, S.S. (1952). On the dual of some balanced incomplete block designs. Biometrics, 8, 66-72.

53. Tukey, J. W. (1951). Quick and dirty methods in statistics. Part II. Simple analysis for standard designs. Proc. Fifth Annual Convention, Am. Soc. for Quality Control, 189-197.

54. Wald, A. (1943). On the efficient design of statistical investigations. Ann. Math. Stat., 14, 134-140.

55. Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. Jour. Agri. Sci., 26. 424-455.

56. Yates, F. (1937). A further note on the arrangement of variety trials: Quasi-Latin squares. Ann. Eug., 7, 319-332.

Let their initial genotypic frequency vector be

$$(2) \qquad f^{(0)}=\left(f_{22}^{(0)}, f_{21}^{(0)}, f_{20}^{(0)}, f_{12}^{(0)}, f_{11c}^{(0)}, f_{11r}^{(0)}, f_{10}^{(0)}, f_{02}^{(0)}, f_{01}^{(0)}, f_{00}^{(0)}\right)'$$

with vector $f^{(n)}$ defined in an analogous manner corresponding to the nth generation, where the components of these vectors add up to one and where the numbers in the subscript stand for the numbers of genes A or B present in the corresponding genotypes. Also the letters c and r stand for the coupling and the recessive phases of the double heterozygotes. The prime in (2) denotes the transpose of a matrix.

We consider first the case where the above population is undergoing self-fertilization indefinitely. The immediate problem is to find the genotypic distribution vector $f^{(n)}$ of the population at the nth stage of continued selfing. To this end, we visualise a discrete time Markov Chain $[X_n\ ;\ n=0, 1, 2,\ldots]$ with the state space S consisting of the ten possible genotypes 1, 2. ..., 10 as listed in (1). The chain is considered to be in state i at nth step if a randomly chosen genotype out of the population (assumed to be infinite in size) at the nth generation turns out to be of type i with $i=1, 2,\ldots,10$. The initial probability distribution is governed by the vector $f^{(0)}$. The Markov property of the chain is obvious, since the genotypic distribution at $(n+1)$th step (generation) depends only on the distribution at the nth step. The only thing left in order to specify the Markov chain completely is the one-step transition probability matrix (P_{ij}) which is stationary in the present case. This is obtained by considering for each genotype individually the genotypic composition it produces after it is once selfed. For instance, given $X_n=5$ (*i.e.* AB/ab), the vector of probabilities $P_{5j}=P_r(X_{n+1}=j|X_n=5)\ ; j=1, 2,\ldots,10$, is given by

$$(q^2/4, pq/2, p^2/4, pq/2, q^2/2, p^2/2, pq/2, p^2/4, pq/2, q^2/4),$$

where $q=1-p$, and so on. The matrix $P=(P_{ij})$ obtained in this manner is given by

$$(3) \qquad P=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ q^2/4 & pq/2 & p^2/4 & pq/2 & q^2/2 & p^2/2 & pq/2 & p^2/4 & pq/2 & q^2/4 \\ p^2/4 & pq/2 & q^2/4 & pq/2 & p^2/2 & q^2/2 & pq/2 & q^2/4 & pq/2 & p^2/4 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus following the usual Markov chain argument the genotypic distribution in the nth generation is given by

$$(4) \qquad f^{(n)}=(P')^n.f^{(0)}.$$

The problem now is to find the nth power of P needed in (4). For this, on solving the characteristic equation $| P-\lambda I | =0$ of matrix P for λ, we observe that the ten eigen values of P are given by

$$\left[1, 1, 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1-2pq}{2}, \frac{1-2p}{2} \right]. \tag{5}$$

Fortunately for the present case the matrix P lends itself to a spectral representation. Following the standard approach (see Karlin[3]) for finding the spectral representation of a matrix by obtaining the left and right eigen vectors for various eigen values, we find that

$$P=MDL \tag{6}$$

where

(7) $D=dg(1, 1, 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, a_2, a_3)$,

$$(8)\ M=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ a_1/2 & pa_1 & pa_1 & a_1/2 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ pa_1 & a_1/2 & a_1/2 & pa_1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$(9) L=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 1/2\sqrt{2} & -1/2\sqrt{2} & 1/2\sqrt{2} & -1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} & -1/\sqrt{2} & 1/2\sqrt{2} & -1/\sqrt{2} & 1/2\sqrt{2} \\ -a_1a_3/\sqrt{2} & 0 & a_1a_3/\sqrt{2} & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & a_1a_3/\sqrt{2} & 0 & -a_1a_3/\sqrt{2} \end{bmatrix}$$

with $ML=I$ and

$$(10) \qquad a_1=\frac{1}{(1+2p)}\ ; \quad a_2=\frac{1-2pq}{2}\ ; \quad a_3=\frac{1-2p}{2}$$

Here I is a 10×10 identity matrix. Also we have used the convention of writing an $n\times n$ diagonal matrix $A=(a_{ij})$ with $a_{ij}=\delta_{ij}a_{ii}$ by $A=dg(a_{11}, a_{22},\ldots, a_{nn})$. Using (6) and the fact that $ML=I$, we finally have

$$(11) \qquad P^n=MD^nL,$$

Yielding for $n=0, 1, 2,\ldots$,

$$(12) \qquad f^{(n)}=L'D^nM'\ f^{(0)},$$

where

$$(13) \qquad D^n=dg\left(1, 1, 1, 1, \frac{1}{2^n}, \frac{1}{2^n}, \frac{1}{2^n}, \frac{1}{2^n}, a_2^n, a_3^n\right).$$

Writing (12) descriptively for later use, we have the expressions for $f_{ij}^{(n)}$ for various i, j values, given in table 1.

It is clear from (3) that the states 1, 3, 8 and 10 are absorption state while the remaining six states are transient. Thus it follows from the theory of finite Markov chains that with probability one the ultimate absorption takes place to one of the four absorption states. The transient states here correspond to those of heterozygotes while the absorption states to those of homozygotes. Thus as expected, the heterozygosity disappears eventually with probability one, with the proportions of various homozygotes in the ultimate population given by letting $n\to\infty$ in $f_{ij}^{(n)}$, yielding

$$(14) \quad \begin{cases} f_{22}^{(\infty)}= f_{22}^{(0)}+\frac{1}{2}\left(f_{21}^{(0)}+f_{12}^{(0)}\right)+\frac{1}{2}\left[a_1 f_{11_c}^{(0)}+(1-a_1) f_{11_r}^{(0)}\right] \\ f_{20}^{(\infty)}= f_{20}^{(0)}+\frac{1}{2}\left(f_{21}^{(0)}+f_{10}^{(0)}\right)+\frac{1}{2}\left[(1-a_1)\ f_{11_c}^{(0)}+a_1\ f_{11_r}^{(0)}\right] \\ f_{02}^{(\infty)}= f_{02}^{(0)}+\frac{1}{2}\left(f_{12}^{(0)}+f_{01}^{(0)}\right)+\frac{1}{2}\left[(1-a_1)\ f_{11_c}^{(0)}+a_1 f_{11_r}^{(0)}\right] \\ f_{00}^{(\infty)}= f_{00}^{(0)}+\frac{1}{2}\left(f_{10}^{(0)}+f_{01}^{(0)}\right)+\frac{1}{2}\left[a_1 f_{11_c}^{(0)}+(1-a_1) f_{11_r}^{(0)}\right], \end{cases}$$

where $f_{ij}^{(\infty)}=0$ for transient states. Furthermore the population approaches homozygosity (fixation) as $n\to\infty$ at least as fast as $(\frac{1}{2})^n$ tends to zero.

2.1. **Distribution of Time to Homozygosity.** Having found that after continuous selfing the population approaches with probability one to homozygosity, it is natural to ask as to how much time it takes before it attains homozygosity with respect to one or the other factor or both. To this end, let T_A and T_B denote the

TABLE 1

Elements of f^(n)

$$f_{22}^{(n)}=f_{22}^{(0)}+\frac{1}{2}\left(f_{12}^{(0)}+f_{21}^{(0)}\right)\left[1-\left(\frac{1}{2}\right)^{n}\right]+\frac{1}{2}f_{11_c}^{(0)}\left[a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}-a_1a_3^{n+1}\right]+\frac{1}{2}f_{11_r}^{(0)}\left[1-a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}+a_1a_3^{n+1}\right]$$

$$f_{21}^{(n)}=\left(\frac{1}{2}\right)^{n}f_{21}^{(0)}+\frac{1}{2}\left(f_{11_c}^{(0)}+f_{11_r}^{(0)}\right)\left[\left(\frac{1}{2}\right)^{n}-a_2^{n}\right]$$

$$f_{20}^{(0)}=f_{20}^{(0)}+\frac{1}{2}\left(f_{21}^{(0)}+f_{10}^{(0)}\right)\left[1-\left(\frac{1}{2}\right)^{n}\right]+\frac{1}{2}f_{11_c}^{(0)}\left[1-a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}+a_1a_3^{n+1}\right]+\frac{1}{2}f_{11_r}^{(0)}\left[a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}-a_1a_3^{n+1}\right]$$

$$f_{12}^{(n)}=\left(\frac{1}{2}\right)^{n}f_{12}^{(0)}+\frac{1}{2}\left(f_{11_c}^{(0)}+f_{11_r}^{(0)}\right)\left[\left(\frac{1}{2}\right)^{n}-a_2^{n}\right]$$

$$f_{11_c}^{(n)}=\frac{1}{2}f_{11_c}^{(0)}\left[a_2^{n}+a_3^{n}\right]+\frac{1}{2}f_{11_r}^{(0)}\left[a_2^{n}-a_3^{n}\right]$$

$$f_{11_r}^{(n)}=\frac{1}{2}f_{11_c}^{(0)}\left[a_2^{n}-a_3^{n}\right]+\frac{1}{2}f_{11_r}^{(0)}\left[a_2^{n}+a_3^{n}\right]$$

$$f_{10}^{(n)}=\left(\frac{1}{2}\right)^{n}f_{10}^{(0)}+\frac{1}{2}\left(f_{11_c}^{(0)}+f_{11_r}^{(0)}\right)\left[\left(\frac{1}{2}\right)^{n}-a_2^{n}\right]$$

$$f_{0}^{(n)}=f_{02}^{(0)}+\frac{1}{2}\left(f_{01}^{(0)}+f_{12}^{(0)}\right)\left[1-\left(\frac{1}{2}\right)^{n}\right]+\frac{1}{2}f_{11_c}^{(0)}\left[1-a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}+a_1a_3^{n+1}\right]+\frac{1}{2}f_{11_r}^{(0)}\left[a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}-a_1a_3^{n+1}\right]$$

$$f_{01}^{(n)}=\left(\frac{1}{2}\right)^{n}f_{01}^{(0)}+\left(f_{11_c}^{(0)}+f_{11_r}^{(0)}\right)\left[\left(\frac{1}{2}\right)^{n}-a_2^{n}\right]$$

$$f_{00}^{(0)}=f_{00}^{(n)}+\frac{1}{2}\left(f_{10}^{(0)}+f_{01}^{(0)}\right)\left[1-\left(\frac{1}{2}\right)^{n}\right]+\frac{1}{2}f_{11_c}^{(0)}\left[a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}-a_1a_3^{n+1}\right]+\frac{1}{2}f_{11_r}^{(0)}\left[1-a_1-\left(\frac{1}{2}\right)^{n}+\frac{1}{2}a_2^{n}+a_1a_3^{n+1}\right]$$

times when the population reaches for the first time homozygous state with respect to gene pairs $A—a$ and $B-b$ respectively. so that

$$(15)\quad P_r(T_A \leqslant m,\ T_B \leqslant n)=P_r(X_m \varepsilon\{1, 2, 3, 8, 9, 10\}) \text{ and } X_n \varepsilon\{1, 3, 4, 7, 8, 10\}).$$

Let $m \leqslant n$. Using the fact that the states 1, 3, 8 and 10 are the absorption states and are homozygous with respect to both factors, we have

$$(16)\quad P_r\left(T_A \leqslant m, T_B \leqslant n \right) = f_{22}^{(m)} + f_{20}^{(m)} + f_{02}^{(m)} + f_{00}^{(m)}$$
$$+ f_{21}^{(m)} P_r(X_n \ \varepsilon\{1, 3\} \mid X_m = 2)$$
$$+ f_{01}^{(m)} P_r(X_n \ \varepsilon\{8, 10\} \mid X_m = 9).$$

From the stationarity property of the Markov Chain it follows that

$$Pr(X_n \varepsilon\{1, 3\} \Big| X_m = 2) = \left(f_{22}^{(n-m)} + f_{20}^{(n-m)} \Big| f_{21}^{(0)} = 1 \right)$$
$$(17)\quad Pr(X_n \ \varepsilon\{8, 10\} \Big| X_m = 9) = \left(f_{02}^{(n-m)} + f_{00}^{(n-m)} \Big| f_{01}^{(0)} = 1 \right)$$

Now on using the expressions for $f_{ij}^{(n)}$ in (17) from table I, we obtain

$$(18)\quad Pr(T_A \leqslant m, T_B \leqslant n) = f_{22}^{(m)} + f_{20}^{(m)} + f_{02}^{(m)} + f_{00}^{(m)} + f_{21}^{(m)} + f_{01}^{(m)}$$
$$- \left(\frac{1}{2}\right)^{n-m} \left(f_{21}^{(m)} + f_{01}^{(m)} \right).$$

Similarly for $m \geqslant n$, we have

$$(19)\quad Pr(T_A \leqslant m, T_B \leqslant n) = f_{22}^{(n)} + f_{20}^{(n)} + f_{02}^{(n)} + f_{00}^{(n)} + f_{12}^{(n)} + f_{10}^{(n)}$$
$$- \left(\frac{1}{2}\right)^{m-n} \left(f_{12}^{(n)} + f_{10}^{(n)} \right).$$

Using (18) and (19) and the expressions of $f_{ij}^{(n)}$ of table I, one easily obtains after some algebra the joint distribution of T_A and T_B as follows :

$$(20)\quad \begin{cases} Pr(T_A = T_B = 0) = f_{22}^{(0)} + f_{20}^{(0)} + f_{02}^{(0)} + f_{00}^{(0)} \\ Pr(T_A = 0, T_B = n) = (\tfrac{1}{2})^n \left(f_{21}^{(0)} + f_{01}^{(0)} \right) ;\ \text{for } n \geqslant 1. \\ Pr(T_A = m, T_B = 0) = (\tfrac{1}{2})^m \left(f_{21}^{(0)} + f_{10}^{(0)} \right) ;\ \text{for } m \geqslant 1. \\ Pr(T_A = T_B = n) = (a_2)^n \left(f_{11c}^{(0)} + f_{11r}^{(0)} \right),\ \text{for } n \geqslant 1. \\ Pr(T_A = m, T_B = n) = (\tfrac{1}{2})^{n-m}\, pq(a_2)^{m-1} \left(f_{11c}^{(0)} + f_{11r}^{(0)} \right) ;\ \text{for } 1 \leqslant m < n, \\ Pr(T_A = m, T_B = n) = (\tfrac{1}{2})^{m-n}\, pq(a_2)^{n-1} \left(f_{11c}^{(0)} + f_{11r}^{(0)} \right) ;\ 1 \leqslant n < m. \end{cases}$$

In the rest of the paper, we shall freely use the convention of replacing a subscript by dot to indicate that a summation has been carried out over that subscript; for instance,

$$f_{\cdot 1}^{(n)}=f_{01}^{(n)}+f_{11c}^{(n)}+f_{11r}^{(n)}+f_{21}^{(n)},$$

$$f_{11.}^{(n)}=f_{11r}^{(n)}+f_{11c}^{(n)};\ \text{and so on.}$$

Now using (20) it is easy to establish that—

$$E(T_A)=2f_{1.}^{(0)}\ ;\ E(T_B)+2f_{\cdot 1}^{(0)},$$

$$\mathrm{Var}(T_{A)}=2f_{1.}^{(0)}\left(3-2f_{1.}^{(0)}\right);\ \mathrm{Var}(T_B)=2f_{\cdot 1}^{(0)}\left(3-2f_{\cdot 1}^{(0)}\right)$$

(21) $$\mathrm{Cov}(T_A,T_B)=\left[\frac{6f_{11.}^{(0)}}{1+2pq}-2f_{1\cdot}^{(0)}f_{.1}^{(0)}\right]$$

$$\rho_{T_A T_A}=\frac{\left[3f_{11.}^{(0)}/(1+2pq)\right]-f_{1.}^{(0)}f_{.1}^{(0)}}{\left[f_{.1}^{(0)}f_{1.}^{(0)}\left(3-2f_{1.}^{(0)}\right)\left(3-2f_{.1}^{(0)}\right)\right]^{\frac{1}{2}}}$$

As expected the correlation coefficient ρ between T_A and T_B depends both on the initial distribution $f^{(0)}$ and on how closely the two factors are linked. The closer are they linked—that is smaller is the value of p—the greater is the correlation between them. In particular if $f_{11.}^{(0)}=1$, then

$$\rho T_A\,T_A=(1-4pq)/(1+2pq),$$

which is non-negative and is equal to 1 or 0 according as $p=0$ or $\frac{1}{2}$.

Consider now another random variable T denoting the time when for the first time the population attains homozygosity with respect to both the factors. One can obtain the distribution of T by using the joint distribution of T_A and T_B given by (20) but more easily by noticing that

(23) $Pr(T\leqslant n)=f_{22}^{(n)}+f_{20}^{(n)}+f_{02}^{(n)}+f_{00}^{(n)}$; $n=0,1,2,\ldots$, so that for $n\geqslant 1$,

(24) $Pr(T=n)=Pr(T\leqslant n)-Pr(T\leqslant n-1)$

$$=(1,0,1,0,0,0,0,1,0,1)(f^{(n)}-f^{(n-1)})$$

$$=(1,0,1,0,0,0,0,1,0,1)L'\,(D^n-D^{n-1})M'f^{(0)}$$

Here the last step follows from (12). Simplifying (24), we have

$$(25)\qquad Pr(T=0)=f_{22}^{(0)}+f_{20}^{(0)}+f_{02}^{(0)}+f_{00}^{(0)}$$

$$Pr(T=n)=(\tfrac{1}{2})^n\left[f_{21}^{(0)}+f_{12}^{(0)}+f_{10}^{(0)}+f_{01}^{(0)}\right]$$

$$+f_{11.}^{(0)}\cdot\left[(\tfrac{1}{2})^{n-1}-\left(\frac{1+2pq}{2}\right)(a_2)^{n-2}\right];\ n\geqslant 1,$$

which yields

$$E(T)=2\left(f_{21}^{(0)}+f_{12}^{(0)}+f_{10}^{(0)}+f_{01}^{(0)}\right)+\frac{2(1+4pq)}{(1+2pq)}f_{11.}^{(0)}$$

$$(26)\quad \text{Var}(T)=6\left(f_{21}^{(0)}+f_{12}^{(0)}+f_{10}^{(0)}+f_{01}^{(0)}\right)+2f_{11.}^{(0)}\left[\frac{3+26pq+24p^2q^2}{(1+2pq)^2}\right]$$

$$-[E(T)]^2.$$

It is clear from (26) that the expected time it takes to attain complete homozygosity increases with p(Note that p takes values between 0 and $\frac{1}{2}$) with minimum value when $p=0$, the complete linkage case and maximum when $p=\frac{1}{2}$, the case where the factors segregate independently. For the special case with $f_{11.}^{(0)}=1$, we have

$$(27)\quad E(T)=\frac{2(1+4pq)}{(1+2pq)}\ ;\ \text{Var}(T)\approx\frac{2(1+10pq-8p^2q^2)}{(1+2pq)^2}.$$

From (27) one finds that behaviour of Var(T) differs from that of $E(T)$ in that it does not always monotonically increase with p. In fact Var (T) is equal to 2 at $p=0$, from where it increases with p until it reaches its maximum value 11/4 at $p=\frac{1}{2}(1-\sqrt{\frac{1}{3}})\approx.211$ after which it decreases to 8/3 at $p=\frac{1}{2}$.

2.2. **Loss of Heterozygosity**. This section deals with the study of loss of heterozygosity for which there appears to be no standard single measure available in literature, particularly when more than one loci are involved. Two of the measures however appear to be more in common use, namly, F^*_n and F_n as defined below for the nth generation.

Following Ghai [2], let $H_0^{(n)}$, $H_1^{(n)}$ and $H_2^{(n)}$ denote the proportion in the nth generation of homozygotes, single heterozygotes and the double heterozygotes respectively, so that

$$H_0^{(n)}+H_1^{(n)}+H_2^{(n)}=1.$$

Let $$H^{(n)}=H_1^{(n)}+H_2^{(n)}$$

denote the proportion of *all* heterozygotes in the *n*th generation. F_n is then defined to be the loss in heterozygosity (single and double heterozygosity combined) at *n*th generation relative to that in the initial population and is given by

$$(28) \qquad F_n = 1 - \frac{H^{(n)}}{H^{(0)}}.$$

Again if Z_n denote the number of heterozygous loci in a genotype randomly drawn from the population at the *n*th generation, we have

$$(29) \qquad P(Z_n = k) = H_k^{(n)} \quad ; \; k = 0, 1, 2.$$

F_n^* is now defined to be the loss in the mean number of heterozygous loci that is $E(Z_n)$ relative to $E(Z_0)$, and is given by

$$(30) \qquad F_n^* = 1 - \frac{E(Z_n)}{E(Z_0)} = 1 - \frac{H_1^{(n)} + 2H_2^{(n)}}{H_1^{(0)} + 2H_2^{(0)}}$$

Another quantity which is of some interest is $\text{Var}(Z_n)$ given by

$$(31) \qquad \text{Var}(Z_n) = \left(H_1^{(n)} + 4H_2^{(n)} \right) - \left(H_1^{(n)} + 2H_2^{(n)} \right)^2$$

Notice that the two measures F_n and F^*_n coincide for the case where only one locus is under consideration. For our present case using the results of Table 1, we have

$$(32) \qquad F_n = \frac{\left(f_{1.}^{(0)} + f_{.1}^{(0)} \right) \left[1 - (\tfrac{1}{2})^n \right] - f_{11.}^{(0)} \left[1 - \left(\frac{1-2pq}{2} \right)^n \right]}{\left(f_{1.}^{(0)} + f_{.1}^{(0)} - f_{11.}^{(0)} \right)},$$

$$(33) \qquad F_n^* = 1 - (\tfrac{1}{2})^n,$$

and

$$E(Z_n) = (\tfrac{1}{2})^n f_{11.}^{(0)}$$

$$(34) \qquad \text{Var}(Z_n) = (\tfrac{1}{2})^n \left[f_{1.}^{(0)} + f_{.1}^{(0)} - (\tfrac{1}{2})^n f_{11.}^{(0)} \right]^2 + 2 \left(\frac{1-2pq}{2} \right)^n f_{11.}^{(0)} .$$

It is interesting to note that whereas the measure F_n depends both on the linkage fraction p and the initial distribution $f^{(0)}$, the measure F_n^* on the other hand is independent of both of these. As expected, both these measures tend to unity as $n \to \infty$. Again, $E(Z_n)$ is independent of p while $\text{Var}(Z_n)$ is not. $\text{Var}(Z_n)$ increases with the decrease in p. These observations concerning Z_n are not new and can be found in a remark made by Kempthorne page (80), [4].

Remark. Ghai [2] has used the measure F_n in order to study the loss in heterozygosity for the case of k independently segregating factors where the population is subjected successively to mixed random mating and selfing. In this case, he has observed that F_n for k independent loci cannot be expected in terms of the value of F_n obtained for a single locus except for the case where the population is completely selfed in successive generations. To this we may now add keeping (32) in mind that this observation still holds in the presence of linkage even if the population is completely selfed. Here in order to obtain F_n for a single locus it is understood that we ignore the other locus completely. Contrary to the behaviour of F_n, the measure F^*_n which turns out to be

$$F^*_n = \frac{v}{2-v}\left[1-\left(\frac{v}{2}\right)^n\right] \tag{35}$$

for the case considered by Ghai [2], is independent of the number of independently segregating factors involved. (Here v is the proportion of the population self-fertilised at each generation and the remaining $(1-v)$ kept under random mating). As such the above observation made for F_n no longer holds if we instead use the measure F^*_n. In this sense then F_n^* may be preferred to F_n, for it does not depend upon the number of independent or linked factors but only on the mode of the mating system used.

2.3. **Some Genetic Properties of the Population Undergoing Selfing.** Let us consider two quantitative genetic characters, one governed by the gene pair A-a and the other by B-b, both linked. In Mather's notation [6], Table 2 gives the various genotypic frequencies in the nth generation along with their genotypic values.

TABLE 2

Genotypic values and Genotypic frequencies

	BB	*Bb*	*bb*	*Total*
AA	$f^{(n)}_{22}$ (d_a, d_b)	$f^{(n)}_{21}$ (d_a, h_b)	$f^{(n)}_{20}$ $(d_a, -d_b)$	$f^{(n)}_{2.}$
Aa	$f^{(n)}_{12}$ $(-h_a, d_b)$	$f^{(n)}_{11.} = f^{(n)}_{11_c} + f^{(n)}_{11_r}$ (h_a, h_b)	$f^{(n)}_{10}$ $(h_a, -d_b)$	$f^{(n)}_{1.}$
aa	$f^{(n)}_{02}$ $(-d_a, d_b)$	$f^{(n)}_{01}$ $(-d_a, h_b)$	$f^{(n)}_{00}$ $(-d_a, -d_b)$	$f^{(n)}_{0.}$
Total	$f^{(n)}_{.2}$	$f^{(n)}_{.1}$	$f^{(n)}_{.0}$	1

Let $M_a^{(n)}$ and $M_b^{(n)}$ be the genotypic means, $\sigma_{aa}^{(n)}$ and $\sigma_{bb}^{(n)}$ the genotypic variance of the two characters respectively. Also, let $\sigma_{ab}^{(n)}$ be their covariance and $\rho_{ab}^{(n)}$ their correlation coefficient. The expressions for these quantities can be easily derived using results of Table I and are given as

$$M_a^{(n)} = d_a \left(2q_A - 1\right) + \frac{h_a}{2^n} f_{1.}^{(0)} \tag{36}$$

$$M_b^{(n)} = d_b \left(2q_B - 1\right) + \frac{h_b}{2^n} f_{.1}^{(0)} \tag{37}$$

$$\sigma_{aa}^{(n)} = d_a^2 \left[4q_A\left(1 - q_A\right) - \frac{1}{2^n} f_{1.}^{(0)}\right] + h_a^2 \frac{f_{1.}^{(0)}}{2^n} \left(1 - \frac{f_{1.}^{(0)}}{2^n}\right) - 2d_a h_a \frac{f_{1.}^{(0)}}{2^n} \left(2q_A - 1\right) \tag{38}$$

$$\sigma_{bb}^{(n)} = d_b^2 \left[4q_B \left(1 - q_B\right) - \frac{1}{2^n} f_{.1}^{(0)}\right] + h_b^2 \frac{f_{.1}^{(0)}}{2^n} \left(1 - \frac{f_{.1}^{(0)}}{2^n}\right) - 2d_b h_b \frac{f_{.1}^{(0)}}{2^n} \left(2q_B - 1\right) \tag{39}$$

$$\sigma_{ab}^{(n)} = \sigma_L^{(n)} + \sigma_P^{(n)} \tag{40}$$

$$\rho_{ab}^{(n)} = \frac{\sigma_L^{(n)}}{\sqrt{\sigma_{aa}^{(n)} \sigma_{bb}^{(n)}}} + \frac{\sigma_P^{(n)}}{\sqrt{\sigma_{aa}^{(n)} \sigma_{bb}^{(n)}}} \tag{41}$$

where

$$\sigma_L^{(n)} = d_a d_b \left(f_{11_c}^{(0)} - f_{11_r}^{(0)}\right)\left(\frac{1-2p}{1+2p}\right)\left[1 - \left(\frac{1-p}{2}\right)^n\right] + f_{11.}^{(0)} \left(\frac{h_a h_b}{2^n}\right) [(1-2pq)^n - (\tfrac{1}{2})^n]. \tag{42}$$

$$\begin{aligned}\sigma_P^{(n)} = {} & d_a d_b \left(f_{22}^{(0)} + f_{00}^{(0)} - f_{20}^{(0)} - f_{02}^{(0)} - 4q_A q_B + 2q_A + 2q_B - 1\right) \\ & + \frac{h_a h_b}{2^{2n}} \left(f_{11.}^{(0)} - f_{1.}^{(0)} f_{.1}^{(0)}\right) + \frac{d_a h_b}{2^n}\left[\left(f_{21}^{(0)} - f_{01}^{(0)}\right) - f_{.1}^{(0)} \left(2q_A - 1\right)\right] \\ & + \frac{d_a d_b}{2^n}\left[\left(f_{12}^{(0)} - f_{10}^{(0)}\right) - f_{1.}^{(0)} (2q_B - 1)\right],\end{aligned} \tag{43}$$

and q_A and q_B are the gene frequencies of genes A and B respectively given by

$$(44) \qquad q_A = f_{2.}^{(n)} + \frac{1}{2} f_{1.}^{(n)} = f_{2.}^{(0)} + \frac{1}{2} f_{1.}^{(0)} \; ; q_B = f_{.2}^{(n)} + \frac{1}{2} f_{.1}^{(n)}$$

$$= f_{.2}^{(0)} + \frac{1}{2} f_{.1}^{(0)} ,$$

keeping in mind that the gene frequencies remain unchanged from generation to generation. It is interesting to note that the first component of the genetic correlation coefficient (41) is entirely due to the presence of linkage and is zero if $p = \frac{1}{2}$. On the other hand the second component of (41) is based only on the initial genotypic distribution $f^{(0)}$ and is independent of p. As expected, the first component disappears again if $f_{11.}^{(0)} = 0$.

One can easily find the limiting expressions for the above quantities as $n \to \infty$. Also one may study the behaviour of these quantities for several special cases such as the one with complete dominance or with no dominance etc. Finally, if instead the same character is governed by both the factors $A-a$ and $B-b$ and if the affects are additive, one can derive the expressions for the genotypic mean and variance in a similar manner. However, we shall not touch these various possibilities here any further. Instead in the next section, we proceed to consider the important case of mixed random mating and selfing.

3. **Population undergoing mixed random mating and self-fertilization.** In this section we consider the case where each generation is produced by subjecting a proportion v of the previous generation to self-fertilization and the remaining portion $u = 1 - v$ to random mating, starting with an initial population with the distribution vector $f^{(0)}$ of (2). Let $r_{11}^{(n)}$, $r_{01}^{(n)}$, $r_{01}^{(n)}$ and $r_{00}^{(n)}$ denote the proportions of gametes AB, Ab, aB and ab respectively produced by the nth generation, so that

$$(45) \qquad \begin{cases} r_{11}^{(n)} = f_{22}^{(n)} + \frac{1}{2}\left(f_{21}^{(n)} + f_{12}^{(n)} + q f_{11c}^{(n)} + p f_{11r}^{(n)}\right) \\ r_{10}^{(n)} = f_{20}^{(n)} + \frac{1}{2}\left(f_{21}^{(n)} + f_{10}^{(n)} + p f_{11c}^{(n)} + q f_{11r}^{(n)}\right) \\ r_{01}^{(n)} = f_{02}^{(n)} + \frac{1}{2}\left(f_{12}^{(n)} + f_{01}^{(n)} + p f_{11c}^{(n)} + q f_{11r}^{(n)}\right) \\ r_{00}^{(n)} = f_{00}^{(n)} + \frac{1}{2}\left(f_{10}^{(n)} + f_{01}^{(n)} + q f_{11c}^{(n)} + p f_{11r}^{(n)}\right) \end{cases}$$

(45) can be rewritten in the matrix form as

$$r^{(n)}=Bf^{(n)}, \tag{46}$$

where

$$r^{(n)}=\left(r_{11}^{(n)},\ r_{10}^{(n)},\ r_{01}^{(n)},\ r_{00}^{(n)}\right), \tag{47}$$

and

$$B=\begin{bmatrix} 1 & \frac{1}{2} & 0 & \frac{1}{2} & \frac{q}{2} & \frac{p}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 1 & 0 & \frac{p}{2} & \frac{q}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{p}{2} & \frac{q}{2} & 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{q}{2} & \frac{p}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 1 \end{bmatrix} \tag{48}$$

For the later developments, we need the following two lemmas.

Lemma 1. A population with genotypic distribution vector $f^{(0)}$ *is in equilibrium with respect to random mating if and only if*

$$f^{(0)}=\left(\left[r_{11}^{(0)}\right]^2,\ 2r_{11}^{(0)}r_{10}^{(0)},\ \left[r_{10}^{(0)}\right]^2,\ 2r_{11}^{(0)}r_{01}^{(0)},\ 2r_{11}^{(0)}r_{00}^{(0)},\ 2r_{10}^{(0)}r_{01}^{(0)},\ 2r_{10}^{(0)}r_{00}^{(0)},\ \left[r_{01}^{(0)}\right]^2,\ 2r_{01}^{(0)}r_{00}^{(0)},\ \left[r_{00}^{(0)}\right]^2\right) \tag{49}$$

and

$$r_{11}^{(0)}r_{00}^{(0)}=r_{10}^{(0)}r_{01}^{(0)}, \tag{50}$$

where $r_{ij}^{(0)}$'s are as defined in (45) for $n=0$.

The proof of this lemma can be found in Kempthorne (pages 38-41, [4]),

Lemma 2. For any distribution vector $f^{(0)}$ *satisfying the conditions of lemma 1.*

$$B(P')^n f^{(0)}=r^{(0)},\ n=0,\ 1,\ 2,\ldots. \tag{51}$$

The proof follows by direct computation and from the facts that $P^n=MD^nL$ and that $r_{11}^{(0)}r_{00}^{(0)}=r_{10}^{(0)}r_{01}^{(0)}$.

From hereon we assume that our initial population is in equilibrium with respect to random mating so that the distribution vector $f^{(0)}$ satisfies the conditions of lemma 1. From (4), (46) and (51), it follows that if this population is subjected successively to complete selfing, then $r^{(n)}=r^{(0)}$ for all n. In our case, this being true separately for both the portions of the population under selfing and under random mating at each generation, we conclude that the gametic frequency vector $r^{(n)}$ remains unchanged over all generations even when the population is subjected simultaneously to both types of mating systems in the manner specified above. Furthermore, from this it follows that under mixed self-fertilization and random mating, the distribution vector of the nth generation is given by

$$(52) \qquad f^{(n)}=uf^{(0)}+vP'f^{(n-1)},$$

where $f^{(0)}$ is as given in (49). Iterating (52) over n and using (11) we obtain

$$(53) \qquad f^{(n)}=uL'[I+vD+v^2D^2+\ldots+v^{n-1}D^{n-1}]M'f^{(0)}+v^nL'D^nM'f^{(0)}.$$

This simplifies further to

$$(54) \qquad f^{(n)}=uL'G_nM'f^{(0)}+v^nL'D^nM'f^{(0)},$$

where

$$(55) \quad G_n=dg\left(\frac{1-v^n}{u}, \frac{1-v^n}{u}, \frac{1-v^n}{u}, \frac{1-v^n}{u} \frac{1-\left(\frac{v}{2}\right)^n}{\left(\frac{1+u}{2}\right)}, \frac{1-\left(\frac{v}{2}\right)^n}{\left(\frac{1+u}{2}\right)},\right.$$

$$\left.\frac{1-\left(\frac{v}{2}\right)^n}{\left(\frac{1+u}{2}\right)} \frac{1-\left(\frac{v}{2}\right)^n}{\left(\frac{1+u}{2}\right)} \frac{1-\left(v\frac{1-2pq}{2}\right)^n}{1-v\left(\frac{1-2pq}{2}\right)} \frac{1-\left(v\frac{1-2p}{2}\right)^n}{1-v\left(\frac{1-2p}{2}\right)}\right).$$

One can immediately find the various elements of $f^{(n)}$ similar to those in Table 1 from (54). Again on letting $n\to\infty$ in (54) we have the limiting genotypic distribution vector given by

$$(56) \qquad f^{(\infty)}=u\,L'\,G\,M'f^{(0)},$$

where

$$(57) \qquad G=dg\left(\frac{1}{u}, \frac{1}{u}, \frac{1}{u}, \frac{1}{u}, \frac{2}{1+u}, \frac{2}{1+u}, \frac{2}{1+u}, \frac{2}{1+u},\right.$$

$$\left.\frac{2}{2-v(1-2pq)}, \frac{2}{2-v(1-2p)}\right)=(I-vD)^{-1}.$$

Using (54) one can as in section 2.3. study the behaviour of various genetic properties of the distribution at nth generation, with changes in u, p and other basic elements. As expected, the expressions of these properties are bit lengthy and we will not tackle them here. We close this section with the final remark that in the presence of random mating the problem of ultimate attainment of homozygosity as discussed in section 2.1 does not arise here.

4. Concluding Remarks. In section 2.2, the author has avoided labelling F_n as the coefficient of inbreeding, simply because he is unaware of any extension of the concept of coefficient of inbreeding to the case of linked factors. In the event there is no such extension available, this appears to be an interesting problem for further investigation. Again the methods used in section 2 based on the study of a Markov chain are similar to what Kempthorne [4] calls as the "Generation matrix method". It is however more revealing to study these problems in the light of theory of Markov chains wherever possible, as some of the already available results of the fairly well developed theory of Markov chains can be readily utilised. The results of section 3 have been generalised to the case where the initial population has an arbitrary distribution and will be communicated elsewhere (see Puri [7]). Also, the above model can be made more realistic by incorporating factors such as selection etc. Unfortunately, however, this makes the algebra somewhat more involved. Finally, I hope that the inadequacies of this paper will not disguise my respects, admiration and affection for Dr. V.G. Panse, to whom it is dedicated.

REFERENCES

1. Ghai, G.L. (1964). The genotypic composition and variability in plant populations under mixed self-fertilization and random mating, *J. Ind. Soc. Agr. Stat.* **16**, pp. 93-125.

2. Ghai, G.L. (1966). Loss of heterozygosity in populations under mixed random mating and selfing, *J. Ind. Soc. Agr. Stat.*, **18**, pp. 71-81

3. Karlin, S. (1966). *A first course in Stochastic Processes*, Academic Press, New York.

4. Kempthorne, O. (1957). *An Introduction to Genetic Statistics*, John Wiley and Sons, Inc., London.

5. Li, C.C. (1967). Genetic equilibrium under selection, *Biometrics*, **23**, pp. 397-484.

6. Mather, K. (1949). *Biometrical Genetics*, Methuen, London.

7. Puri, Prem S. (1967). A model with mixed self-fertilization and random mating in the presence of linkage. *Mimeo Series* No. 135, *Department of Statistics, Purdue University, Lafayette, Indiana.*

ESTIMATION OF POOLED MEAN

By

K. NAGABHUSHANAM*

1. Pooled Mean :

Suppose that we have a sequence of random variables $[X_j]$ with finite expectations $[E(X_j)]=[m_j]$, and with $\underset{N\to\infty}{Lt}\ \frac{1}{N}\sum_{j=1}^{N} m_j=M$. Then we shall refer to M as the pooled mean of the sequence of the random variables.

2. Estimation of the pooled mean.

We wish to estimate M when the variates are not independent. It is clear that some restrictions are needed for the sample mean, $\frac{1}{N}\sum_{j=1}^{N} X_j$, to be consistent estimator of M, despite the intutional feeling that the sample mean would be asymptotically efficient in a fairly wide class of estimators. It is to this question that we address ourselves here. A set of sufficient conditions will be obtained in what follows.

3. Mean square convergence,

If $\underset{N\to\infty}{Lt}\ E\,|\,\xi_N-C\,|^2=0$, we say that the sequence of random variables (ξ_j) converges in the mean square to the constant C. Since mean square convergence implies convergence in probability, one way to seek a set of sufficient conditions for the consistency of estimation is to seek conditions for the mean square convergence. A result embodying such a proposition is known as a mean ergodic theorem. We shall establish such a theorem, assuming that second moments, *viz.* $E(X_j X_s)$ exist finitely.

4. Theorem

If (*i*) for λ in $(-\pi, \pi)$ the sequence of functions :**

$$\sigma_N(\lambda)=\frac{1}{2\pi N}\left[(\pi+\lambda)\sum_{j=1}^{N}\rho(j,j)+\sum_{k=-N+1}^{N-1}{}'\left(\sum_{\substack{j,\,s\\ s-j=k}}\rho(j,s)\,\frac{e^{-ik\lambda}-e^{ik\pi}}{-ik}\right)\right],$$

where $\rho(j,s)=\text{Cov.}(X_j, X_s)$, converges to a function $\sigma(\lambda)$ at all of its continuity points,

(*ii*) $\rho_N(0)=\frac{1}{N}\sum_{j=1}^{N}\rho(j,j)$ are uniformly bounded for all N,

---

*Andhra University.

**The prime above the second sigma on the right side indicates that the value zero is omitted in the summation for k.

and (*iii*) the limit function $\sigma(\lambda)$ is continuous at $\lambda=0$,

then

$$\underset{N\to\infty}{Lt} \; E\left|\frac{1}{N} \sum_{j=1}^{N} X_j - M\right|^2 = 0.$$

Remark:

The function $\sigma(\lambda)$ is just like the spectral function of a wide sense stationary process, and hence may be called the spectrum of the process $[X_j]$. (See K. Nagabhushanam and C.S.K. Bhagavan [1]). The requirement of the continuity of the specturm at the origin is the familiar condition when the process is stationary. It is now seen that the same condition is needed for the generalised spectrum when the random variables do not constitute a stationary process.

Proof.

We shall not go into all the details of the proof, but only outline the main stages in it.

(*a*) For $\lambda_2 > \lambda_1$ in $(-\pi, \pi)$ it is readily verified that

$$\sigma_N(\lambda_2) - \sigma(\lambda_1) = \frac{1}{2\pi N} \int_{\lambda_1}^{\lambda_2} E\left|\sum_{t=1}^{N} eit \;\; X(t) - m(t)\right|^2 d\lambda \geqslant 0,$$

so that $\sigma_N(\lambda)$ is a non-decreasing function of λ for each N.

(*b*) Whatever N may be, $\sigma_N(-\pi) = 0$, and hence $\sigma_N(\lambda)$ is non-negative.

(*c*) If $\rho_N(r) = \frac{1}{N} \sum_{\substack{j,\, s=1 \\ s-j=r}}^{N} \rho(j, s)$,

it can be verified that

$$\rho_N(r) = \int_{-\pi}^{\pi} eir\lambda \;\; d\sigma_N(\lambda).$$

(*d*) From the uniform boundedness of $[\rho_N(0)]$, we get that each $\sigma_N(\lambda)$ as well as $\sigma(\lambda)$ is a bounded measure function on $(-\pi, \pi)$ with a total measure less than a fixed constant.

(*e*) The saltus positions of a bounded measure function are at most denumerably infinite in number, so that we can choose $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ as small as desired and also such that $-\varepsilon_1$ and ε_2 are continuity points of the spectral function $\sigma(\lambda)$.

(*f*) $\dfrac{\left(\sin \dfrac{2N-1}{2}\lambda\right)\Big/\left(\dfrac{2N-1}{2}\lambda\right)}{\left(\sin\dfrac{\lambda}{2}\right)\Big/\left(\dfrac{\lambda}{2}\right)}$ and its limit for $\lambda=0$ are less than or equal to unity.

$$(h)\ E\left|\frac{1}{N}\sum_{j=1}^{N}(X_j-m_j)\right|^2=\frac{1}{N^2}\sum_{k=-N+1}^{N-1}\sum_{\substack{r,s=1\\ s-r=k}}^{N}\rho(r,s)$$

$$=\frac{1}{N}\sum_{k=-N+1}^{N-1}\rho_N(k)$$

$$=\int_{-\pi}^{\pi}\frac{1}{N}\ e^{-i(N-1)\lambda}\left(\sum_{r=0}^{2N-2}e^{ir\lambda}\right)d\sigma_N(\lambda)$$

$$=\frac{1}{N}\int_{\pi}^{\pi}e^{-i(N-1)\lambda}\frac{\sin\dfrac{2N-1}{2}\lambda}{\sin\dfrac{\lambda}{2}}d\sigma_N(\lambda)$$

$$\leqslant\frac{1}{N}\int_{-\pi}^{\pi}\frac{\left|\sin\dfrac{2N-1}{2}\lambda\right|}{\left|\sin\dfrac{\lambda}{2}\right|}d\sigma_N(\lambda)$$

$$=I_1+I_2+I_3,$$

where $$I_1=\frac{1}{N}\int_{-\pi}^{-\varepsilon_1}\frac{\left|\sin\dfrac{2N-1}{2}\lambda\right|}{\left|\sin\dfrac{\lambda}{2}\right|}d\sigma_N(\lambda).$$

$$I_2=\frac{2N-1}{N}\int_{-\varepsilon_1}^{\varepsilon_2}\frac{\left|\left(\sin\dfrac{2N-1}{2}\lambda\right)\Big/\dfrac{2N-1}{2}\lambda\right|}{\left|\sin\dfrac{\lambda}{2}\Big/\dfrac{\lambda}{2}\right|}d\sigma_N(\lambda),$$

and $$I_3=\frac{1}{N}\int_{\varepsilon_2}^{\pi}\frac{\left|\sin\dfrac{2N-1}{2}\lambda\right|}{\left|\sin\dfrac{\lambda}{2}\right|}d\sigma_N(\lambda).$$

SOLUTIONS TO SOME FUNCTIONAL EQUATIONS AND THEIR APPLICATIONS TO CHARACTERISATION OF PROBABILITY DISTRIBUTIONS

by

C.G. KHATRI AND C. RADHAKRISHNA RAO*

Summary. Three sets of results are contained in this paper.

The first is on a new matrix product. If A and B are two matrices of orders $p \times r$ and $q \times r$ respectively, and if $\alpha_1, \ldots, \alpha_r$ are column vectors of A and $\beta_1, \ldots, \beta_r$ are those of B then the new product $A \odot B$ is the partitioned matrix

$$(\alpha_1(\times)\beta_1 : \alpha_2(\times)\beta_2 \vdots \ldots \vdots \alpha_r(\times)\beta_r$$

where $(\times)$ denotes the Kronecker product. Propositions involving the new product of matrices are stated.

The second is on the solution of functional equations of two types. One is of the form

$$\sum_{u=1}^{p} c_{ju}\psi(e_u't) + \sum_{i=1}^{r} b_{ji}\phi_i(\alpha_i't) = g_j \text{ (constant)}$$
$$j = 1, \ldots, q$$

involving a vector variable t where e_u are unit vectors of an identity matrix of order p, α_i are given column vectors and ψ_u, ϕ_i are unknown continuous functions.

Another is of the form

$$\sum_{j=1}^{n} d_{ij}\phi(b_jt) = g_i, \quad i = 1, \ldots, q$$

involving an unknown function ϕ of a single variable t. Conditions under which the unknown functions in these two types of equations are polynomials of an assigned degree are given.

The third, on the characterisation of normal and gamma distributions, extends the earlier work of the authors (Rao, 1967 and Khatri and Rao, 1967) We consider two sets of functions $L_1, \ldots, L_q$ and $M_1, \ldots, M_p$ of independent random variables $X_1, \ldots, X_n$ with the condition

$$E(L_i/M_1, \ldots, M_p) = g_i \text{ (constant)}$$

for $i = 1, \ldots, q$ When L_i and M_j are linear, the X_i have normal distributions When L_i are linear in the reciprocals of the variables and M_j are linear in the variables, the X_i have gamma or conjugate gamma distributions. When the X_i variables are non-negative, L_i are linear in the variables and M_j are linear in the logarithms of the variables, the X_i have gamma distributions. These results are proved under some

---

*Indian Statistical Institute.

(*i*) For sufficiently small ε_1 and ε_2 and sufficiently large N, we can make I_2 as small as desired. Once ε_1 and ε_2 are chosen, we can make each of I_1 and I_3 as small as desired by taking N sufficiently large. From this it follows easily that

$$E \left| \frac{1}{N} \sum_{j=1}^{N} X_j - M \right|^2 = E \left| \frac{1}{N} \sum_{j=1}^{N} (X_j - m_j) + \left(\frac{1}{N} \sum_{j=1}^{N} m_i - M \right) \right|^2$$

can be made as small as desired, which proves the theorem.

REFERENCE

1. K. Nagabhushanam and C.S.K. Bhagavan. Non-stationary Processes and Spectrum (to be published in the Canadian Jour. of Math.).

condition on the compounding coefficients for $p>1$ and in the case of $p=1$ with the further condition that the X_i are identically distributed.

1. **Introduction.** Linnik (1964) considered a functional equation in two variables t_1, t_2 of the type

$$\phi_1(t_1+b_1t_2)+\ldots+\phi_r(t_1+b_rt_2)=\xi_1(t_1)+\xi_2(t_2) \qquad \ldots(1.1)$$

defined for $|t_1|<\partial$, $|t_2|<\partial$, for some $\partial>0$, where $\phi_1,\ldots,\phi_r$ and ξ_1, ξ_2 are unknown continuous functions, and showed, by an extremely elegant method, that all the functions involved in (1.1) must be polynomials provided only that $b_1,\ldots,b_r$ are all different. In a recent paper Rao (1966) considered a slightly extended form of (1.1)

$$\phi_1(t_1+b_1t_2)+\ldots+\phi_r(t_1+b_rt_2)=\xi_1(t_1)+\xi_2(t_2)+Q(t_1,\, t_2) \qquad \ldots(1\ 2)$$

defined for $|t_1|<\partial$, $|t_2|<\partial$, where Q is a quadratic function in t_1, t_2 and showed that each function involved in (1.2) is a polynomial of degree not more than max $(2, r)$ provided that $b_1,\ldots b_r$ are different. In the case of (1.1), without the quadratic function, the degree of each polynomial is found to be utmost r.

We now consider a functional equation in $p>2$ variables $t_1,\ldots, t_p$ of the type

$$\phi_1(\alpha_1'l)+\ldots+\phi_r(\alpha_r'l)=\xi_1(t_1)+\ldots\xi_p(t_p) \qquad \ldots(1.3)$$

defined for $|t_i|<\delta$, $i=1,\ldots, p$, where t represents the column vector of variables $t_1,\ldots,t_p$ and $\alpha_1,\ldots\alpha_r$ are given column vectors. Our object is to determine the conditions on $\alpha_1\ldots\alpha_r$ under which each function in (1.3) is a polynomial and to find an upper bound to the maximum degree of the polynomials. It is shown that more precise estimates of the maximum degree than in the case (1.2) can be found depending on the nature of the vectors $\alpha_1,\ldots,\alpha_r$. The case where the maximum degree is utmost unity (see lemma 4) is of special interest and is considered in some detail. Conditions under which the maximum degree is $k<r$ are given in lemma 5. Thus, an increase in the number of variables in Linnik's equation (1.1) places a restriction on the degree of the polynomials.

As a generalisation of the equation (1.3), we consider multiple equations of the form

$$\sum_{u=1}^{p} c_{ju}\Psi_u(e_4't)+\sum_{i=1}^{r} b_{ji}\phi_i(\alpha_i't)=g_j \qquad \ldots(1.4)$$

defined for $|t_i|<\partial,\ldots$, $i=1,\ldots p$, where $\Psi_1,\ldots,\Psi_p$; $\phi_1,\ldots,\phi_r$ are unknown continuous functions, g_j are constants, e_u are unit column vectors of the identity matrix I_p of order p and $\alpha_1,\ldots,\alpha_r$ are given column vectors. In lemmas 6, 7 and 8, we determine the conditions under which the functions involved in (1.4) are polynomials and under which the maximum degree does not exceed a given number.

Finally, we consider multiple equations of the form

$$\sum_{j=1}^{n} d_{ij}\phi(b_jt)=g_i \text{ (constant)} \qquad \ldots(1.5)$$

$$i=1,\ldots,q$$

in a single variable t defined for $|t| < \partial$, where ϕ is an unknown function. This is a generalisation of the single equation

$$a_1\phi(b_1t)+\ldots+a_n\phi(b_nt)=0 \qquad \ldots(1.6)$$

considered by Rao (1967). It is shown that when $\phi(t)$ is of the form of $c+t\,\psi(t)$ where $\psi(t) \to$ a constant as $t \to 0$, then $\phi(t)$ is a linear function under some conditions on the coefficients.

We use the solutions of the equations (1·3). (1·4) and (1·5) in characterising normal and gamma distributions. These results extend those obtaided in earlier papers by Rao (1967) and Khatri and Rao (1967).

In section 2 of the paper we define a new product of matrices and consider its properties. The solutions of the functional equations (1·3), (1·4) and (1·5) are discussed in section 3 and the main theorems on characterisation of the normal and the gamma distributions are given in sections 4 and 5.

2. A new product of matrices

Let $A=(a_{ij})$ and B be any two matrices. Then the Kronecker product $A(\bar{X})$ B is defined by

$$A(\times)B=(a_{ij}\,B). \qquad \ldots(2\cdot1)$$

If A is $p\times q$ matrix and B is $m\times n$ matrix, then the order of $A(\times)\,B$ is $pm\times qn$.

Now we shall consider two matrices A of order $p\times r$ and B of order $q\times r$ and denote the column vectors of A by $\alpha_1,\ldots,\alpha_r$ and those of B by $\beta_1,\ldots\beta_r$.

Definition : The new product $A\odot B$ is defined by the partitioned matrix

$$A\odot B=(\alpha_1(\times)\beta_1 \,\vdots\, \alpha_2(\times)\beta_2 : \ldots..\alpha_r(\times)\beta_r) \qquad \ldots(2\cdot2)$$

which is of order $pq\times r$.

We state some propositions involving the new product of matrices, which follow from the definition or which can be easily established.

(*i*) It is easy to see that if C is of order $s\times r$ with column vectors $\gamma_1,\ldots\gamma_r$, then

$$A\odot B\odot C=(\alpha_1(\times)\beta_1(\times)\gamma_1 \,\vdots\ldots\vdots\, \alpha_r(\times)\beta_r(\times)\gamma_r) \qquad \ldots(2\cdot3)$$

is of order $pqs\times r$ and

$$(A\odot B)\odot C=A\odot(B\odot C) \qquad \ldots(2\cdot4)$$

and so on.

Further $A\odot B$ and $B\odot A$ differ only in a permutation of rows. Hence the six possible orders of multiplying three matrices A, B, C, lead to matrices which differ only in a permutation of rows.

(*ii*) Let T_1 be a matrix of order $m\times p$ and T_2 of order $n\times q$. Then

$$(T_1(\times)T_2)\,(A\odot B)=T_1A\odot T_2B \qquad \ldots(2\cdot5)$$

(*iii*) If $\alpha_1,\ldots,\alpha_r$ and $\beta_1,\ldots,\beta_r$ are all nonnull vectors, then $A\odot B$ has no null column. If A has a null column vector, then the corresponding column vector in $A \odot B$ is null. Conversely if $A \odot B$ has a null column vector, then the cerresponding column vector in A or B must be null.

(*iv*) If two nonnull columns in $A \odot B$ are proportional, then the two corresponding nonnull column vectors in A as well as in B will be proportional and conversely.

(*v*) Let all the column vectors of B corresponding to independent column vectors of A be nonnull. Then rank $(A \odot B) \geqslant$ rank A. Similarly, if all the column vectors of A corresponding to independent column vectors of B are nonnull, then rank $(A \odot B) \geqslant$ rank B.

(*vi*) If rank $(A \odot B)=r$ and the i_1-th, i_2-th,..., i_u-th column vectors of B are proportional, then the i_1-th, i_2-th,..., i_u-th column vectors of A are linearly independent ; and all column vectors of A and B are nonnull vectors.

(*vii*) If rank $A=r$ which is the number of columns of A and s is the number cf null column vectors in B, then rank $(A \odot B)=r-s$.

Definition. Let $A^{\#}$ be the matrix obtained from $A \odot A$ by deleting the p rows involving the square terms (*i.e.*, by deleting the 1st, $(p+2)$-th,...p^2-th rows), where A is of order $p \times r$.

(*viii*) If rank $A^{\#}=r$, then

(*a*) no two columns of A are dependent, and

(*b*) each column of A contains at least two non-zero elements.

Note. We observe that while rank $(A \odot A) \geqslant$ rank A, it is not possible to make a general statement regarding the relative magnitude of the ranks of A and $A^{\#}$. We give some examples to show that rank $A^{\#}$ may be less than, greater than or equal to rank A.

Consider the matrices

$$A_1=\begin{bmatrix} 1 & \cdot & \cdot & 1 \\ -1 & \cdot & \cdot & 1 \\ \cdot & 1 & 1 & \cdot \\ \cdot & -1 & 1 & \cdot \end{bmatrix}, \quad A_2=\begin{bmatrix} 1 & 1 & 2 & \cdot \\ 1 & \cdot & 1 & 1 \\ \cdot & 1 & 1 & -1 \\ 1 & \cdot & 1 & 1 \end{bmatrix}$$

$$A_3=\begin{bmatrix} 1 & 1 & 2 \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{bmatrix}, \quad A_4=\begin{bmatrix} 1 & 1 & -1 \\ \cdot & 1 & 1 \\ 1 & \cdot & \cdot \end{bmatrix}, \quad A_5=\begin{bmatrix} 1 & 1 & -1 \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{bmatrix}$$

By actual computations we find

(*a*) rank $A_1=4$. rank $A_1^{\#}=2$, and rank $(A_1 \odot A_1^{\#})=4$,

(*b*) rank $A_2=2$, rank $A_2^{\#}=3$, and rank $(A_2 \odot A_2^{\#})=4$,

(c) rank $A_2=2$, rank $A_3^{\#}=3$,

(d) rank $A_4=3$, rank $A_4^{\#}=2$.

(e) rank $A_5=3$, rank $A_5^{\#}=3$.

Definition. Let us denote, for any positive integer s,

$$(A\odot)^s A^{\#}=(A\odot)^{s-1} A\odot A)^{\#}$$

$$=\underbrace{A\odot A\odot \dots A\odot A}_{s \text{ times}}{}^{\#} \qquad \dots(2{\cdot}6)$$

(ix) *If no two column vectors of* $(A\odot)^s A^{\#}$ or $A^{\#}(\odot A)^s$ are proportional, then

(a) no two column vectors of A are proportional, and

(b) each column vector of A has at least two non-zero entries.

(x) Rank $(A\odot)^s A^{\#} \geqslant$ Rank $(A\odot)^t A^{\#}$ for $s \geqslant t \geqslant O$.

(xi) Rank of $A\odot A \geqslant$ rank A where A is of order $p\times r$, but if no two column vectors of A are proportional to each other, then rank of $A \odot A \geqslant$ min. (r, $1+$rank A).

3. Solutions to some functional equations

First we quote a lemma proved in an earlier paper (lemma 2 in Rao, 1966) which is used in proving the main results of this section.

Lemma 1. Let A be $p\times r$ matrix *such that the i-th column vector of A* is not a multiple of any other column vector of A or of any column vector of B of order $p\times m$, and the first element of the *i-th* column vector of A is non-zero (without loss of generality). Then there exists a $2\times p$ matrix.

$$H=\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & h_2 & & h_p \end{bmatrix} \qquad (3.1)$$

such that the matrices

$$C_1=HA,\ C_2=HB \qquad (3.2)$$

of orders $2\times r$ and $2\times m$ respectively satisfy the property that the *i-th* column vector of C_1 *is not a multiple* of any other column vector of C_1 or of any column vector of C_2.

3.1. Functional equation

Consider the functional equation

$$\phi_1(\alpha_i' t)+\dots+\phi_r(\alpha'_r t)=\xi_1(t_1)+\dots+\xi_p(t_p) \qquad (3.3)$$

defined for $|\ t_i\ | < \partial, i=1,\dots p$, where t is a column vector of the variables $t_1, \dots, t_p$ and $\alpha_1, \dots \alpha_r$ are the column vectors of a given matrix A (of order $p\times r$). The functions $\phi_1, \dots, \phi_r, \xi_1, \dots, \xi_p$ are unknown except that they are continuous. The object is to determine the form of these functions under different conditions on the elements of A.

Lemma 2. Let α_i, the *i-th* column vector of A, be not proportional to any other column of A or to any column of I_p an identity matrix of order p. Then the function ϕ_i is a polynomial of maximum degree r.

Proof. Without loss of generality we take the first element of *i-th* column of A as nonzero. By lemma 1, there exists a matrix

$$H=\begin{bmatrix} 1 & 0 \;\dots\; 0 \\ 0 & h_2 \dots h_p \end{bmatrix} \tag{3.4}$$

such that the *i-th* column of $B=HA$ is not proportional to any other column of HA or to any column of H. Let

$$t'=(t_1, \dots, t_p)=(u_1, u_2)H \tag{3.5}$$

Then the equation (3.3) becomes

$$\begin{aligned} &\phi_1(b_{11}u_1+b_{12}u_2)+\dots+\phi_r(b_{r1}u_1+b_{r2}u_2) \\ &=\xi_1(u_1)+\xi_2(h_2u_2)+\dots+\xi_p(h_pu_2) \\ &=\xi_1(u_1)+\eta(u_2) \end{aligned} \tag{3.6}$$

valid for some interval round the origin of u_1 and u_2. In (3.6), (b_{i1}, b_{i2}) is not proportional to (b_{j1}, b_{j2}), $j\neq i$ or to (1, 0) or to (0, 1). Hence the term $\phi_i(b_{i1}u_1+b_{i2}u_2)$ cannot be combined with any other ϕ_j. Then by using Linnik's lemma as stated by Rao (1966), ϕ_i is a polynomial of maximum degree r.

Lemma 3. If no column of A is proportional to any other column of A or to any column of I_p,then $\phi_1, \dots\phi_r$ and $\xi_1, \dots, \xi_p$ are all polynomials of maximum degree r.

Proof. By lemma 2, $\phi_1, \dots, \phi_r$ are all polynomials of maximum degree r, and hence $\xi_1, \dots, \xi_p$ are all polynomials of maximum degree r.

Lemma 4 Consider the matrix $A^{\#}$ of order $p(p-1)\times r$ defined in section 2. If rank $A^{\#}$ is r, then $\phi_1, \dots, \phi_r$ and $\xi_1, \dots, \xi_p$ are all linear functions.

Proof. The proof consists of two parts. By (*viii*) of section 2, rank $A^{\#}=r$ implies that no column of A is a multiple of any other column of A or of any column of I_p. Hence *using lemma 3, all ϕ_i and all ξ_i are polynomials of maximum* degree r.

Now let

$$\begin{aligned} &\phi_i(u)=\lambda_{ir}u^r+..\ \lambda_{i1}u+\lambda_{io},\ i=1,\dots,r, \\ &\xi_j(u)=\mu_{jr}u^r+\dots+\mu_{j1}u+\mu_{jo},\ j=1,\ \dots,\ p \end{aligned} \tag{3.7}$$

and denote $\lambda'_i=(\lambda_{i1},\dots\lambda_{ir})$. Using the functional forms (3.7) in (3.3) and collecting the coefficients of t_it_j, $i\neq j$ we find

$$A^{\#}\lambda_2= \tag{3.8}$$

which implies that $\lambda_2=0$, since $A^{\#}$ has full rank equal to r by assumption. Thus the second degree terms in the polynomial forms (3.7) are absent.

Now collecting the coefficients of $t_i^{\pi_1}\ t_j^{\pi_2}\ t_k^{\pi_3}$, $i\neq j\neq k$ and $(\pi_1+\pi_2+\pi_3)=3$

with at least two π's nonzero, we find

$$(A \odot A^{\#})\lambda_3 = 0$$

or $$(A^{\#} \odot \mathrm{A})\lambda_3 = 0 \tag{3.9}$$

By (x) of section 2, rank $(A \odot A^{\#}) = s$ since rank $A^{\#} = s$. Thus $\lambda_3 = 0$, or the third degree terms are absent. Similarly collecting coefficients of $\underset{i}{t}^{\pi_1} \underset{j}{t}^{\pi_2} \underset{k}{t}^{\pi_3} \underset{l}{t}^{\pi_4}$, $i \neq j \neq k \neq l$ and $(\pi_1 + \pi_2 + \pi_3 + \pi_4) = 4$, with atleast two π's non-zero, we find

$$[(A \odot)^2 A^{\#}]\lambda_4 = 0 \qquad or \qquad [A^{\#}(\odot A)^2]\lambda_4 = 0 \tag{3·10}$$

and by the same argument used to show $\lambda_3 = 0$ we have $\lambda_4 = 0$ and so on. Thus, all terms of degree higher than one are absent in the polynomials (3.7) which proves that $\phi_1, \ldots, \phi_r$ can utmost be of degree one and so must be $\xi_1, \ldots \xi_p$.

Lemma 5. Consider a non-negative integer $s < r-1$. If rank $[(A \odot)^s A^{\#}] = r$, then $\phi_1, \ldots, \phi_r$, $\xi_1, \ldots \xi_p$ are all polynomials of degrce $(s+1)$ utmost.

Proof. Since rank $(A \odot)^s A^{\#} = r$, no two columns of $(A \odot)^s A^{\#}$ are proportional. Hence using (ix) of section 2, no two columns of A are proportional and each column of A has atleast two non-zero entries. Then lemma 2 shows that the functions $\phi_1, \ldots \phi_r$, $\xi_1 \ldots, \S_p$ are all polynomials of maximum degree r.

Since rank $(A \odot)^s A^{\#} = r$, by arguments similar to those of lemma 4, the $(s+2)$-th degree terms in the polynomials are absent. Further the condition rank $(A \odot)^s A^{\#} = r = >$ rank $[(A \odot)^{s+1} \; A^{\#}] = r$. Then the $(s+3)$-th degree terms are absent and so on, so that the maximum degree of $\phi_1 \ldots\ldots \phi_r$ can be $(s+1)$ utmost. This proves lemma 5.

Corollary. Let A be of rank $r (\leqslant p)$ such that each column has atleast two non-zero entries. Then $\phi_1, \ldots\ldots \phi_r$, $\xi_1 \ldots\ldots\ldots \xi_p$ are all quadratic functions.

Proof. Note that in this case $A^{\#}$ has all the column vectors non-null. Hence rank $(A \odot A^{\#}) \geqslant$ rank $A = r$ which is true cnly if rank $(A \odot A^{\#}) = r$. Then by lemma 5, we get the result.

3.2. Functional Equation II

Consider a partitioned matrix

$$\begin{bmatrix} \underset{(q \times p)}{C} & \underset{(q \times r)}{B} \\ \\ \underset{(p \times p)}{I} & \underset{(p \times r)}{A} \end{bmatrix} \tag{...(3.11)}$$

and represent the i-th column vector of A by α_i and the (u, i)-th element of A by a_{ui}. Similarly β_i, b_{ji}, γ_u, c_{ju} are defined for the matrices B and C respectively. The column vectors of I_p are denoted by $e_1 \ldots\ldots e_p$.

Consider the q equations in p unknown $t'=(t_1 \ldots\ldots t_p)$

$$\sum_{u=1}^{p} c_{ju}\, \psi_u\, (e'_u t) + \sum_{i=1}^{r} b_{ji}\, \phi_i(\alpha_1'\, t) = g_j \text{ (constant)} \qquad \ldots(3.12)$$

$$j=1, \ldots\ldots, q$$

defined for $|\, t_i \,| < \delta$, $i=1, \ldots\ldots, p$, where $\psi_1, \ldots\ldots, \psi_p\, ;\ \phi_1, \ldots\ldots, \phi_r$ are continuous functions.

Lemma 6. Let

(*a*) each column of C and B has atleast one non-zero entry, and

(*b*) each column of A is not proportional to any other column of A or to any column of I_p.

Then $\psi_1, \ldots\ldots, \psi_p$ and $\phi_1, \ldots\ldots, \phi_r$ are all polynomials of degree r utmost.

Proof. The proof follows on the same lines as those of lemmas 2 and 3. The condition (*b*) of lemma 6 can be replaced by the more general condition (*b'*).

(*b'*) *Suppose that a column* α_{i_1} *of* A *is proportional to other columns* α_{i_2}, α_{i_3} , ……, of A and some e_m. Then it should be possible to find constant $a_1, \ldots\ldots, a_q$ such that in the equation

$$\sum_j a_j \left[\sum_u c_{ju} \psi_u(e_u' t) + \sum_i b_{ji}\, \phi_i\, (\alpha_i'\, t) - g_j \right] = 0 \qquad \ldots(3.13)$$

the coefficients of functions involving the arguments $\alpha'_{i_2}\ t$, $\alpha'_{i_2}\ t$, $\alpha'_{i_3}\ t$, ……, $e'_m\, t$ are all zero and the coefficient of the function involving the argument $\alpha'_{i_1}\ t$ is not zero. Observe that the equation (3·13) is obtained from the equations (3.12) by multiplying the j-th equation by a_j and adding over j.

Lemma 7. Let in (3 12) rank $(B \odot A^{\#}) = r$. Then $\psi_1, \ldots\ldots, \psi_p, \phi_1, \ldots\ldots\, \phi_r$ are linear functions.

Lemma 8. Let in (3·13), rank $[B \odot (A \odot)^s\, A^{\#}] = r$ (where $s < r-1$). Then $\psi_1, \ldots\ldots, \psi_p, \phi_1, \ldots\ldots, \phi_r$ are polynomials of degree $(s+1)$ utmost.

Notice that on account of (*vi*) and (*viii*) or (*ix*) of section 2, the condition (*b'*) of lemma 6 will be satisfied. Hence, proofs of lemmas 7 and 8 are similar to those of lemmas 4 and 5.

3·3. Functional Equation III

Consider the q equations involving an unknown function ϕ and a single variable t

$$\sum_{j=1}^{n} d_{ij}\, \phi(b_j t) = gi, \quad i=1, \ldots\ldots, q \qquad \ldots(3.13)$$

defined for $|\, t \,| < \delta$, where $b_1, \ldots\ldots, b_n$ are different without loss of generality. By multiplying the i-th equation of (3.13) by a'_i and adding over i, we obtain a compound equation

$$a_1 \phi(b_1\, t) + \ldots\ldots\ldots + a_n \phi(b_n\, t) = h$$

where

$$\sigma_j = \sum_i a_i' \, d_{ji}, \; h = \sum_i a_i' g_i \qquad \ldots(3.14)$$

Lemma 9. Let there exist constants $a_1', \ldots\ldots, a_q'$ such that the coefficients $a_1, \ldots\ldots a_q$ satisfy the following conditions :—

(*a*) $\Sigma \, a_i b_i = 0$, and

(*b*) if $a_1, \ldots\ldots, a_s (s \leqslant n)$ are non-zero without loss of generality, then there is only onc element in the set $(| b_1 |, \ldots, | b_s |)$ which exceeds the others. If $| b_1 | > \max (| b_2 |, \ldots\ldots, | b_s |)$, without loss of generality, then $a_i b_i$, $i = 2, \ldots\ldots, s$ have the same sign but different from that of $a_1 b_1$. Further let $\phi(t) = c + t\psi(t)$ where $\psi(t) \to$ constant as $t \to 0$. Then $\phi(t)$ is a linear function of t.

We observe that, if some of the b_i are the same we can rewrite the equations (3.13) by combining some of the terms such that in the resulting equations the b_i are all different though with a lesser number of terms. The conditions of the lemma can then be stated in terms of coefficients of the reduced equations.

The proof is similar to that of lemma 2 given by Rao (1967), using the compound equation (3.14).

4. Characterization of the Normal Law

Let $X_1, \ldots\ldots, X_n$ be the independent random variables, not necessarily identically distributed. Consider a linear function

$$a_1 X_1 + \ldots\ldots + a_n X_n \qquad \ldots(4.1)$$

with all non-zero coefficients, which by suitable scaling can be written as

$$L = X_1 + \ldots\ldots + X_n. \qquad \ldots(4.2)$$

Further let

$$\begin{array}{l} b_{11} X_1 + \ldots + b_{1n} X_n \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_{p1} X_1 + \ldots + b_{pn} X_n \end{array} \qquad \ldots(4.3)$$

be p independent linear functions, which by a suitable transformation can be written in a canonical form

$$\begin{array}{l} M_1 = X_1 + c_{11} X_{p+1} + \ldots + c_{1, n-p} X_n \\ \cdot \quad \cdot \quad \cdot \quad \cdots \quad \cdot \\ M_p = X_p + c_{p1} X_{p+1} + \ldots + c_{p, n-p} X_n. \end{array} \qquad \ldots(4.5)$$

Denote the matrix of the c_{ji} coefficients in the equation (4.5) by C which is of order $p \times (n-p)$ and let c_i be the *i-th* column vector of C.

Theorem 1. Let $p > 1$ and each column vector of C be not proportional to any other column vector of C or to a column of the identity matrix. Further let $X_1, \ldots, X_n$ have finite first moments. Then the condition

$$E(L \mid M_1, \ldots, M_p) = 0 \qquad (4.6)$$

is necessary and sufficient that $X_1, \ldots X_n$ are all normally distributed.

Proof. The condition (4.6) is equivalent to

$$E(Le^{it_1M_1+\ldots+it_pM_p})=0 \qquad \ldots(4.7)$$

which gives the functional equation

$$\psi_1(t_1)+\ldots+\psi_p(t_p)+\psi_{p+1}(c_1't)+\ldots+\psi_n(c'_{n-p}t)=0 \qquad \ldots(4.8)$$

valid for $| t_i | <\delta$, $i=1,\ldots,p$, where $\psi_i=\phi_i'/\phi_i$, ϕ_i being the characteristic function of X_i. Using lemma 2, we find each ϕ_i is a polynomial and hence X_i is normal.

Theorem 2. Suppose that c_i is proportional to $c_{i1}, c_{i2},\ldots$ and to e_s, the s-th column of a unit matrix. Let

$$c_{i1}=\lambda_{i1}c_i,\ c_{i2}=\lambda_{i2}c_i,\ldots,e_s=\lambda_s c_i. \qquad \ldots(4.9)$$

If $\lambda_{i1}, \lambda_{i2},\ldots,\lambda_s$ are all of the same sign and atleast one column of C contains two non-zero entries, then X_i, X_{i1}, $X_{i2},\ldots$ are all normally distributed when (4.6) holds and the first moments of the variables exist.

The proof follows on the same lines as in Theorem 3 of Rao (1967).

We now consider two sets of q and $p\geqslant 2$ linear functions which are all independent and hence can be written in canonical forms

$$L_j=\sum_{u=1}^{p} c_{ju}X_u+\sum_{i=p+1}^{n} b_{j,\ t-p}X_t,\ \ j=1,\ldots,q \qquad \ldots(4.10)$$

$$M_i=X_i+\sum_{u=p+1}^{n} a_{iu-p}X_u,\ i=1,\ 2,\ldots,p$$

and examine the restrictions on the coefficients under which the conditions

$$E(L_j/M_1,\ldots,M_p)=0, \quad j=1,\ldots,q \qquad \ldots(4.11)$$

imply normality of the variables $X_1,\ldots,X_n$.

Theorem 3. Let $A=(a_{ij})$ be of order $p\times n-p$, $B=(b_{ij})$ be of order $q\times n-p$ and $C=(c_{ij})$ be of order $q\times p$. Then under the restrictions on the elements of A, B, C mentioned in lemmas 6, the codition (4.11) and the existence of the first moments of the variables imply normality of the variables $X_1,\ldots,X_n$.

Proof. The result follows from the functional equations obtained from the conditions

$$E(L_j\ e^{it_1M_1+\ldots+it_pM_p})=0, \quad j=1,\ldots,q \qquad \ldots(4.12)$$

by applying lemmas 6.

The result is, however, true under more general situations than those considered in lemmas 6. When some of the vectors in the matrix A are proportional, a theorem similar to Theorem 2 could be stated.

The case of $p=1$ needs special discussion which we state in Theorem 4.

Theorem 4. Let $X_1,\ldots,X_n$ be independent and identically distributed random variables, and

$$L_i=\sum_{j=1}^{n} d_{ij}X_j,\ \ i=1,\ldots,q \qquad \ldots(4.13)$$

be q linearly independent functions and

$$M=b_1X_1+\ldots+b_nX_n \qquad \ldots(4.14)$$

be a function linearly independent of $L_1,\ldots,L_q$. Then under the conditions of lemma 9 on the coefficients b_i and d_{ij}, the conditions

$$E(L_i/M)=0, \quad i=1,\ldots,q \qquad \ldots(4.15)$$

imply that X_i is normally distributed.

Proof. Let $\phi(t)=f'(t)/f(t)$ where $f(t)$ is the characteristic function of X_i. Then the condition (4.15) gives

$$\Sigma \ d_{ij}\phi(b_jt)=0, \quad i=1,\ldots,q \qquad \ldots(4.16)$$

valid for $|\ t\ |<\delta$. It is shown by R.N. Pillai* (in personal correspondence) that the condition (4.16) implies that $E(X_i^2)<\infty$, so that ϕ is of the form $t\psi(t)$ where $\psi(t)\rightarrow$ constant as $t\rightarrow 0$. Then applying lemma 9, we find ϕ is a linear function and hence X_i is normally distributed.

5. Characterisation of the Gamma Distribution

Let $X_1,\ldots,X_n$ be independent random variables, not necessarily identically distributed. We consider two situations where X_i are non-negative and when X_i are arbitrary. When X_i are non-negative, let

$$Y_i=\log X_i$$

$$M_i=a_{i1}Y_1+\ldots+a_{in}Y_n \qquad \ldots(5.1)$$

and when X_i are arbitrary, let

$$M_i'=a_{i1}X_1+\ldots+a_{in}X_n \qquad \ldots(5.1')$$

In a previous paper (Khatri and Rao, 1967), it was proved, under some conditions, that

$$E(\sum_{j=1}^{n} b_{ij}e^{\ Y_j} \Big/ M_1,\ldots,M_{n-q})=g_i, \text{ (constant)},$$

$$i=1,\ldots,q\leqslant n-2 \qquad \ldots(5.2)$$

when X_i are non-negative or

$$E\Big(\Sigma\, b_{ij}\ X_j^{-1} \Big/ M_1',\ldots,M_{n-q}') = g_i', \text{ (constant)},$$

$$i=1,\ldots,q\leqslant(n-2) \qquad \ldots(5.2')$$

when X_i are arbitrary, implies that $X_1,\ldots,X_n$ have gamma or conjugate gamma distributions. Now, we consider the conditions under which

$$E(\ \Sigma\ b_{ij}e^{Y_j}\ /\ M_1\ldots,M_p)=g_i, \text{ (constant)},$$

$$i=1,\ldots,\ q\ ;\ q+p\leqslant n \qquad \ldots(5.3)$$

where X_i are non-negative, or

$$E\left(\ b_{ij}\ X_j^{-1} \Big/ M_1' \ ,\ldots,\ M_p' \ \right)=g_i', \quad \text{(constant)}$$

$$i=1,\ldots,\ q\ ;\ q+p\leqslant n \qquad \ldots(5.3')$$

---

*The paper by Pilllai containing this result will be published elsewhere.

when X_i are arbitrary, implies that $X_1,\ldots, X_n$ have gamma or conjugate gamma distributions. It may be noted that the generalization consists in reducing the conditioning variables to p from the full complement of $n-q$ considered in the earlier paper by the authors (Khatri and Rao, 1967).

We observe that the linear functions $M_1,\ldots,M_p$ can be considered to be linearly independent in which case they may be represented in the canonical form

$$\begin{aligned} M_1 &= Y_1 + a_{11} Y_{k+1} + \ldots + a_{1,\, n-p}\, Y_n \\ &\ldots \quad \ldots \quad \ldots \quad \ldots \\ M_p &= Y_p + a_{p1} Y_{p+1} + \ldots + a_{p,\, n-p}\, Y_n \end{aligned} \qquad \ldots(5.4)$$

The functions $M_1',\ldots M_p'$ have a similar representation. Let A denote the matrix (a_{ij}). We shall first consider the case of $q=1$.

Theorem 5. Let $Y_1,\ldots, Y_n$; $M_1,\ldots, M_p$ and $M_1', \ldots, M_p'$ be as considered in (5.1), (5.1′) and (5.4) respectively. Further let $1<p\leqslant(n-1)$ and the rank of $A^{\#}$ defined in section 2 is $r=(n-p)$. Then the condition

$$E(e^{Y_1} + \ldots + e^{Y_n} / M_1,\ldots,M_p) = g \text{ (constant).} \qquad \ldots(5.5')$$

when X_i are non-negative and $E(X_i)<\infty$ for all i, implies that $X_1,\ldots, X_n$ have gamma distributions. The condition

$$E\left(X_1^{-1} + \ldots + X_n^{-1} / M_1', \ldots, M_p' \right) = g \text{ constant} \qquad \ldots(5.5')$$

when X_i are arbitrary and $E\left(X_i^{-1} \right) \neq 0$ and $\left| E\left(X_i^{-1} \right) \right| < \infty$ for all i, implies that $X_1,\ldots, X_n$ have gamma distributions when $E(X_i) > 0$ and conjugate gamma distribution when $E(X_i) < 0$.

Proof. Let

$$\phi_j(t) = \left[\int e^{Y_j} e^{itY_j}\, dF(Y_j) \right] \div \left[\int e^{itY_j}\, dF(Y_j) \right] \qquad \ldots(5.6)$$

when X_i are non-negative, or

$$\phi_j(t) = \left[\int x_j^{-1} e^{itx_j} dF(x_j) \right] \div \left[\int e^{itx_j} dF(x_j) \right] \qquad \ldots(5.6')$$

when X_i are arbitrary and $E\left(X_i^{-1}\right) \neq 0$. Then the condition (5.5) or (5.5′) is equivalent to

$$\phi(t_1) + \ldots + \phi_p(t_p) + \phi_{p+1}\left(\alpha_1' \, t \right) + \ldots + \phi_n\left(\alpha_{n-p}' \, t \right) = \text{constant} \qquad \ldots(5.7)_t$$

valid for $| t_i | < \delta$, $i=1,\ldots, p$, where α_i is the i *th* column vector of A and $t'=(t_1\ldots t_p)$. We now apply lemma 4 which shows that, under the condition rank $A^{\#} = r = n-p$, the functions $\phi_1,\ldots,\phi_n$ are all linear in t. In such a case, it is shown in the earlier paper of Khatri and Rao that X_i has a gamma distribution for each i.

It may be noted that in the conditions (5.5) and (5.5′) we could have chosen the more general function under the expectation,

$$a_1 e^{Y_1}+,\ldots,+a_n e^{Y_n},\ a_i \neq 0,\ i=1,\ldots,n \qquad \ldots(5.8)$$

or

$$a_1 X_1^{-1}+\ldots+a_n X_n^{-1},\ a_i \neq 0,\ i=1,\ldots,n \qquad \ldots(5.8')$$

and obtain the same result.

Theorem 6. Let $X_1,\ldots,X_n$ be independent and non-negative random variables with finite expectations and $Y_1,\ldots,Y_n$ be as defined in (5·1). Consider

$$L_i=c_{i1}e^{Y_1}+\ldots+c_{ip}\,e^{Y_p}+b_{i1}\,e^{Y_{p+1}}+\ldots+b_{in-p}\,e^{Y_n} \qquad \ldots(5.9)$$
$$i=1,\ldots,q$$

$$M_j=Y_j+a_{i1}Y_{p+1}+\ldots+a_{jn-p}Y_n$$
$$j=1,\ \ldots,\ p>1. \qquad (5.10)$$

If the matrices $C=(c_{ij})$, $B=(b_{ij})$ and $A=(a_{ij})$ satisfy the conditions of lemma 7, and

$$E(L_i/M_1,\ldots,M_p)=g_i \text{ (constant)}$$
$$i=1,\ldots,\ q, \qquad (5.11)$$

then $X_1,\ldots X_n$ have gamma distributions.

Proof. It is seen that the condition (5.11) gives rise to the functional equation of the form (3.12) and hence an application of lemma **7** yields the desired result.

Theorem 7. Let $X_1\ldots,X_n$ be independent variables with non-zero and finite expectations for $X_1^{-1},\ldots,X_n^{-1}$, and

$$L_i=c_{i1}X_1^{-1}+\ \ldots\ +c_{ip}\,X_p^{-1}+b_{i1}\,X_{p+1}^{-1}+\ldots+b_{in-p}\,X_n^{-1}$$
$$i=1,\ \ldots,\ q. \qquad (5.15)$$
$$M'_j=X_j+a_{j1}X_{p+1}+\ldots+a_{jn-p}X_n$$
$$j=1,\ldots,p>1. \qquad (5.16)$$

If the matrices $C=(c_{ij})$, $B=(b_{ij})$ and $A=(a_{ij})$ satisfy the conditions of lemma 7, and

$$\mathrm{E}(\mathrm{L}_i \mid M'_1,\ldots,M'_p)=g_i \text{ (constant)}$$
$$i=1,\ldots,q \qquad (5.17)$$

then X_i has a gamma or a conjugate gamma distribution according as $E(X_i)>0$ or <0, $i=1,\ldots,n$.

Theorem 8. Let $X_i,\ldots,X_n$ be non-negative independent and identically distributed variables. Consider

$$L_i=\Sigma d_{ij}\,e^{Y_j}, \qquad i=1,\ldots.q \qquad (5.12)$$
$$M=b_1Y_1+\ldots+b_nY_n \qquad (5.13)$$

where the coefficients b_i and d_{ij} satisfy the conditions of lemma 9. Then the conditions

$$E(L_i \mid M)=g_i\,(constant),\ i=1,\ldots,q \qquad (5.14)$$

and $E(X_i \log X_i)$ is bounded imply that X_i has gamma distribution.

Theorem 9. Let $X_1,\ldots,X_n$ be independent and identical variables (not necessarily non-negative) such that $E(1/X_i)$ exists and is non-zero. Further let

$$L'_i=\Sigma d_{ij} X^{-1},\ i=1,\ldots,q$$

and

$$M'=b_1X_1+\ldots+b_nX_n$$

where the coefficients b_i and d_{ij} satisfy the conditions of lemma 9. Then the conditions

$$E(L'_i/M')=g'_i\ \text{(constant)},\ i=1,\ldots,q$$

imply that X_i has a gamma or a conjugate gamma distribution according as $E(X_i)>0$ or <0.

The proofs of Theorems 8 and 9 follow on the same lines as the other theorems.

REFERENCES

1. Khatri, C.G. and Rao, C. Radhakrishna (1967) : Some characterisations of the gamma distribution. *Tech. Report. No. 23/67 of Indian Satistical Institute, Calcutta-35.*
2. Rao, C. Radhakrishna (1967) : On some characterisations of the Normal law. *Sankhya,* 29. 1-14.
3. Rao, C. Radhakrishna (1966) : Characterisations of the distribution of random variables in linear structural relations. *Sankhya,* 28, 251-260.
4. Yu. V. Linnik (1964) : *Decomposition of Probability distribution,* Oliver and Boyd.

AGRICULTURAL CENSUS AND THE NEW STRATEGY FOR AGRICULTURAL DEVELOPMENT

By

J.S. SARMA

Comprehensive and fairly reliable agricultural statistics giving area and production of crops, land utilisation and irrigation are available in India, for quite some time; but essentially these are aggregative data, with the Survey Number (or *Khasra* Number) as the primary unit of enumeration. The primary data are aggregated at the successive geographical levels of administrative units, such as village, tehsil and district in all the States except West Bengal, Orissa and Kerala. The yield statistics in these States as well as both area and yield statistics in the States of West Bengal, Orissa and Kerala are available at the district level. For meeting the growing needs of planning in Agriculture, data are needed *with the cultivators' holding* as the unit as this is the basic unit for decision-making. Agricultural production is the result of the decisions and actions taken by the farmer regarding what to produce, when to produce, and how to produce on his holding with the resources that are available to him or that can be augmented. What the Government can do is to create the basic environment and facilities for the farmers' increasing their production and provide the necessary resources—technical know-how, material inputs and finance. A knowledge of the detailed structure of agricultural holdings and the resources available is thus essential for the success of agricultural planning.

Agricultural Census provides for the collection of data on the *structure* of agriculture, covering information on the following items :

(*i*) Number and area of agricultural holdings and their principal characteristics such as size, type, form of tenure, utilization of land, etc.

(*ii*) Number and characteristics of farm population including their employment.

(*iii*) Area under crops and number of livestock and poultry.

(*iv*) Agricultural machinery and availability of transport facilities, and

(*v*) Irrigation, drainage and the use of fertilisers and soil dressings.

While information on the above mentioned items is to be collected and presented holding-wise, data on volume of production of *principal agricultural products* could be presented for aggregates of areas.

It will be usefull to draw attention at this stage to the distinction between the concepts of *operational* and ownership holdings. A cultivator's holding may be defined as "all land that is used wholly or partly for agricultural production and is cultivated alone or with the assistance of others without regard to ownership, size or location." This definition of the holding for the purpose of Agricultural Census refers to an operational concept. A holding may consist of two or more parcels of land even if widely separated, provided they form part of the same operational unit. The holding includes all cultivated land irrespective of whether a particular crop is grown on that area or not. On the other hand, an ownership holding is all agricultural land that is owned individually or jointly with others by an owner (or a group of owners) irrespective of whether it is cultivated by him or not. Thus, the operational holding of a cultivator consists of all agricultural land (irrespective of location) owned by him *minus* land leased out by him to others *plus* land taken on lease by him from others for cultivation purposes.

Earlier World Agricultural Censuses

The First World Agricultural Census sponsored by the Food and Agriculture Organisation of the United Nations was organised in different countries generally in the year 1950. It was designed essentially to provide, on a comparable basis, information on the number and characteristics of agricultural holdings and of the people who secure their livelihood from agriculture, areas under crops and numbers of livestock, and the volume of production of important agricultural products. It was, in principle, conceived as a direct enumeration of individual holdings in each country, although it was anticipated that other means of obtaining the desired information might be preferable or necessary in some areas or countries.

The question of India's participation in the First World Agricultural Census was considered in detail by a Teachnical Committee set up by the Government of India under the Chairmanship or W.R. Natu. This Committee of which Dr. Panse was also a member, recommended that the opportunity provided by the World Agricultural Census should be utilised for placing the agricuitural statistics of the country on a firm footing. The Committee's Report entitled : "Coordination of Agricultural Statistics in India" (September 1949) gave a blueprint for improvement of agricultural statistics for the country and indicated at the same time the detailed steps necessary for obtaining the information desired by the FAO. However, in India, the data required for World Agricultural Census were obtained on the basis of a sample survey organised as part of the Eighth Round of the National Sample Survey (August 1954—April 1955).

The Second World Agricultural Census was sponsored by the FAO in the year 1960. This time too, the data in respect of India were collected through a sample survey organised as part of the NSS in its 16th and 17th Rounds. The next (the third) World Agricultural Census is due in 1970. The Government of India have agreed to participate in the Census, although no firm decision has yet been taken regarding the method to be adopted for the Census, whether the data are to be collected on the basis of a sample survey to be organised as a part of the NSS as in the past, or through complete enumeration on the lines suggested by the Technical Committee or by organising a separate sample survey.

With this background of the Agricultural Census, it will be useful to consider the manner in which data regarding cultivators' operational holdings would be useful for agricultural planning at the different stages of Plan formulation, implementation and evaluation.

The Government of India have launched the New Strategy for Agricultural Development in the year 1966 67 under which concerted efforts are being made to attain self-sufficiency by 1970-71. The main features of the New Strategy are :

(*i*) Bringing, by 1970-71, an area of 32.5 million acres under the High-Yielding Varieties Programme covering exotic varieties of rice, Mexican dwarf varieties of wheat, hybrid varieties of jowar, bajra and maize in irrigated or assured rainfall areas,

(*ii*) Extension of multiple-cropping over 30 million acres of irrigated area and undertaking intensive cultivation measures,

(*iii*) Change in the concept of irrigation from drought-protection to intensive crop production,

(*iv*) Assurance of adequate supplies of fertilisers, pesticides, etc.,

(*v*) Ensuring timely agricultural credit for short, medium and long term periods,

(*vi*) Introduction of improved agricultural implements with a view to improving the efficiency of agricultural operations,

(*vii*) Assurance of remunerative prices to the farmer and measures designed to increase his income through better marketing and storage facilities,

(*viii*) Removal of the disincentives to production caused by outmoded systems of tenure and burdensome tenancies,

(*ix*) Organisation of better extension services through demonstrations and farmers' education etc.,

(*x*) Undertaking special area development programmes in an integrated manner, covering land levelling, preparation for intensive irrigated cultivation, better communications, markets, etc.

High-Yielding Varieties Programme

As against an area of 4·8 million acres reported to have been brought under High-Yielding Varieties Programme in 1966-67, the target for the current year is 15 million acres which is being further raised to 21 million acres in 1968-69. In order to enable planning and execution of the programme in the subsequent years, as also in the subsequent Plan periods, it is important to know which types of farmers have adopted the High-Yielding Varieties Programme ; whether the innovators come entirely from the bigger and medium farm groups or whether the small farmers also have taken up the programme ; whether the farmers put the entire area of their holdings under these varieties or only a part of it ; whether they have put the entire irrigated area under the high-yielding varieties or only a part thereof. Once answers to these questions are obtained, the information could be used in planning the programme for the future, For this purpose, it is necessary to have data on the number of holdings of each size, together with the data regarding area under foodgrains and total cultivated area with further break-downs regarding irrigated and unirrigated areas.

Multiple—Cropping Programme

It is proposed to cover 7·5 million acres under the multiple cropping programme during 1967-68, as against the target of 30 million acres by 1970-71. The measures include replacement of the existing long-duration varieties of crops by those of short-duration which depends upon the extent to which irrigation facilities for the second icrop are assured and the more intensive use of these facilities is made. It is necessary to draw up suitable crop rotations for broad soil types within each local area, taking into account the facilities available by way of irrigation, etc. Here, again, before detailed targets are fixed for each area and programmes are drawn up, information regarding the existing pattern of cropping and intensity of cropping, classified according to different sizes of holdings, is necessary.

The intensity of use of crop land has been found to vary with the size of the holdings and their other characteristics, *e.g.* extent of human and cattle labour used, number and types of irrigation sources, modes of irrigation and the equipment used in irrigation Different cropping intensities and cropping patterns have to be recommended for achieving the maximum returns. Holding-wise data on cropping intensities and patterns in relation to nature of irrigation facilities available or to be developed, and human labour and cattle and mechanical power available have, therefore, to be built up.

Irrigation Programmes

The relationship between irrigation and size of holding needs close study. At present, in many areas, irrigation even from canals, is not perennial and protective irrigation is provided for one crop only. With the popularisation of minor irrigation programmes like those for installation of tubewells, pumpsets, etc. which provide greater manoeuvrability of supply of water and assure more frequent and intensive use of water, programmes for intensive cultivation are receiving a fillip. It is necessary to see what types of crops are grown at present under irrigated conditions in different sizes of holdings, whether the growth of commercial-versus-foodgrain crops bears any relationship to the size of the holding, and whether the differences in types of irrigation sources and combination thereof to supplement each other make much material differences in the intensity and pattern of cropping on holdings of different sizes. Unless the present practices and patterns are studied carefully in relation to farm sizes, it will be difficult to lay down detailed programmes at the field level.

Fertilisers

The consumption of chemical fertilisers has gone up rapidly from about 200 thousand tonnes of N in 1960-61 to 550 thousand tonnes in 1965-66. It is expected to go up to 1150 thousand tonnes in 1967-68. There is considerable demand for fertilisers largely as a resuls of the High Yielding Varieties Programme, the Package Programme (Intensive Agricultural District Programme), Multiple Cropping Programme and the programmes for cultivation of cash crops. Further extension in the distribution of fertilisers requires solutions to such problems as statisfactory arrangements for marketing, credit, etc. To come to meaningful decisions regarding each of these matters, information regarding the size and characteristics of holdings on which chemical fertilisers are at present in use and the progress and problems of adoption of fertilisers on holdings of different sizes, is necessary. Are chemical fertilisers being used intensively by the big farmers, the medium farmers, or, are the small farmers also using these ? If not, what are their specific problems ? Are the different farmers able to meet the requirements of funds from their own resources or have they to depend entirely or partially on external resources ? Are the subsistence farmers also using fertilisers and, if so, how are their credit needs being met ? Information on such items can be obtained at successive points of time and the programmes for fertiliser distribution reoriented from time to time after detailed benchmark information relating to size and characteristics of the holdings becomes available.

Agricultutal Credit

For planning agricultural credit programmes, it is necessary to have information regarding the requirements of funds for different purposes of the different types of holdings. Although these requirements could be worked out on per acre basis taking into account total expenditure, the extent to which these funds are provided by the farmer from his own resources depends on the characteristics of the holding. The bigger farmers are able to invest more in land : they can also meet a greater proportion of their requirements from their own funds. Thus, holdingwise information will help considerably in the implementation of credit programmes.

Agricultural Machinery and Implements

The demand for agricultural machinery and improved agricultural implements is growing rapidly under the impact of the new strategy. Not only are farmers investing more money in minor irrigation works like pumpsets and tubewells, but more and more farmers are also purchasing tractors, power tillers, etc. It would be necessary to know which types of farmers are going in for these machines. Mechanisation is also being resorted to meet the demands for peak period labour under intensive and multiple cropping programmes. In fact to cope with the pressure of agricultural operations connected with the harvesting of the first crop and the sowing of the second crop in the limited time available, the farmer has to resort to partial mechanisation. The extent to which mechanisation can supplement or replace bullock labour needs detailed study. Even small farmers are going in for tractor-hiring to facilitate their cultivation operations. For all these programmes, holdingwise data are necessary.

Incomes snd Prices

The new technology has increased the gross returns of the farmer. It is expected that even a small holding of three acres which is fully irrigated, can have surpluses in terms of agricultural commodities and also become an economic holding, if two irrigated crops of high yielding varities are grown in a year with intensive application of irrigation. If, on the other hand, irrigation is not available or is not assured and the intensity of cropping is less, probably a larger holding would be necessary for giving a reasonable income to the farmer. The size of the economic holding in unirrigated areas with a single-crop pattern will definitely be larger. Different types of data are required at two stages, firstly, for determining for each local situation, the size of the economic holding and secondly, after determining this limit for formulating programmes and policies situated to the different holdings in relation to the economic holding.

For formulating policies of marketing, including procurement, it would be necessary to know the types of farmers who contribute to marketable surpluses. Also, in judging the impact of price support measures, it would be necessary to find out the marketable surpluses of different sizes of farms and the extent the price acts as an incentive to increased production of the small farmers who produce mainly for their own consumption. Thus in evaluating price support measures in agriculture, it will be necessary to see whether the benefits of these measures will accrue to only a few large scale operators or the small farmers also.

Land Reforms

Land reform measures also are an essential pre-requisite for the success of the New Strategy. The farmer should be assured that a major proportion of the extra benefits received from the land go to him. A tenant farmer may not be enthused to put in more efforts if, after putting in extra investment and receiving additional output, he has to part with a larger proportion as rent to the owner. For gauging the magnitude of the problem and for evolving appropriate policy measures, it is necessary to know the distribution of holdings according to tenure and tenancy in each of the areas particularly in the areas selected for the implementation of the High Yielding Varieties Programme.

Extension Services

The country is covered with the national extension service under the Community Development Programme. Under this programme, there are ten village level workers in a Block, each worker being put in charge of ten villages. This coverage has been found to be inadequate for intensive cultivation programmes such as the Package Programme and the High Yielding Varieties Programme. The number and jurisdiction of extension workers to be employed depend on the number and distribution of holdings and their principal characteristi.s, for any given programme. An idea of the characteristics of the farm population will also be useful in formulating for each local area the extension techniques suitable for that area.

Special Area Programmes

It has been experienced that to obtain maximum results in agricultural development, the approach has to be an integrated development of the area. This approach is essential in the areas newly brought under irrigation. Several development measures are necessary such as land levelling, provision of field channels and drainage, conversion from dry to wet cultivation, provision of communications, opening of new markets. For proper planning of these various programmes, data on structure and characteristics of holdings at the time of undertaking the programme is essential. This would indicate the number of farmers coming into the programme, and give an idea of their resources which could form the starting point for reorganisation under irrigated agriculture.

Levels at which data are needed

Thus, it will be seen that for formulating and implementing the programmes under the new strategy, holdingwise information is necessary. In the absence of this information, some targets are fixed at the State level and these are broken up districtwise on pro-rata basis or on some a-priori considerations with the resulting experience of inadequate performance, non-fulfilment of the targets in some areas despite the provision of inputs and lack of supply of inputs in other areas where there is demand. There is no *doubt that had this information been available for each planning unit at* the Block and district levels, performance in Indian agriculture could have vastly improved, through better planning of allocation of resources and through better implementation of the programmes.

In organising the collection of data on a holdingwise basis, an important question that arises is the unit for which these data are required ; for, this has a bearing on the question whether the data should be collected on a complete-enumeration basis or on a sample survey basis. A sample survey would provide holdingwise information at the most at the State level. It can also give data at the district level, but to provide reliable data at this level in respect of holdings sub-divided into various characteristics such as size, tenure, irrigation, type of crop grown (*i.e.* foodgrains or commercial crops) the size of the sample will have to be unduly large, and the cost is likely to be prohibitive. If the data are required at the Block level, complete enumeration seems to be the only practical method. For most of the Agricultural Programmes, the district and the blocks are the planning units. The IADP (Package Programme) is a district programme but the unit of planning is the block. For the high-yielding varieties programme also, certain districts have been selected and the programme is to be implemented in the selected blocks growing the crops under conditions of irrigation or assured rainfall. Thus, the balance of advantage seems to be in favour of having the data on a complete enumeration basis. This would also provide bench-mark data which could be used in organising further sample surveys from time to time, on broader aspects of the holdings.

The way in which agricultural census data can be used in planning and implementation of the programmes may need some elaboration. As already pointed out, the Census provides data as at a point of time or for an agricultural year, basic data regarding the structure and characteristics of agricultural holdings. It may not provide all the data regarding the various relations between the development measures proposed and the holdings. But once these relationships are established on the basis of sample surveys or intensive studies, these could be used in planning the programmes, in conjunction with the basic data on the holdings provided by the census for each local area.

The criticism that in India agricultural planning is imposed from the top and the targets are built up not from below seems to be correct. There is an increasing realisation that, to be successful, the planning has to be done from below. For doing this, however, there are two essential requisites. The first is the availability of data. The second is the availability of trained personnel who could use the data and formulate the relevant programmes. Once the data are available, it should be possible to organise the analysis and interpretation of the data which the planner could use for drawing up detailed plans for smaller geographical areas and aggregating them into district and State Plans.

Limitations of the Census Data

It may be argued that detailed holdingwise information for smaller areas is required not at the stage of planning but at the stage of implementation. This argument may be valid if we still believe in planning at the national level without an eye to the problems of implementation by the numerous decision-making small farms. This apart, any large gap between the period to which the data employed in planning at the macro level relate and the actual period of implementation would render the whole Plan unrealistic and difficult of implementation and achieving the targetted goals. On this ground, some people do not favour a simultaneous agricultural census all over the place and would prefer local inventories being taken up as and when the programmes are drawn up. It may, however, be stated that the main characteristics of holdings do not change in too short a period. There is every advantage in undertaking an agricultural census all over the country with comparable data for different States and regions at a single point of time.

Another shortcoming in the census data to which attention is usually drawn is the fact that these data relate to a year. Weather and climatic conditions which influence agriculture greatly, show considerable annual variations. If the reference period for the census happens to be one of the abnormal years, the data collected would also not be representative. This criticism is, to some extent, true. But the only way out is to postpone the reference period for the census, if prior knowledge regarding the abnormality of the year chosen for census is available.

Other Advantages of the Census

The opportunity provided by the Agricultural Census can be utilised to set the entire structure of agricultural statistics on a sound footing. This occasion would provide an opportunity to the reporting authorities at the primary level to devote sufficient and detailed attention to the reliability and accuracy of data while recording the various agricultural statistics. The supervisory officers can also go into the data carefully and see that the concepts and definitions are adopted uniformly and that the classification has been done properly. In this sense, the agricultural statistics data collected through the Census year can provide a bench-mark for future purposes. It provides an opportunity to bring upto-date complete and rectify the basic records which might have been ignored under the pressure of work on the primary reporter arising from work of a more urgent and immediate nature. The uniformity of concepts and definitions can also be more rigorously enforced. The basis for new types of data can be laid and the primary reporters and the supervisory staff made aware of the changing needs of data in the context of planning for a richer and more diversified agriculture.

Agricultural Census results also provide useful source material for national income statistics or for making further refinements therein. Additional data on prices of agricultural inputs and outputs, livestock products and depreciation on fixed assets are needed for this purpose. Thus, once in ten years the bench-mark data for national income could be made reliable and also internationally comparable by adopting uniform concepts and definitions.

Method of Enquiry to be Adopted

Part of the information pertaining to Agricultural Census can be obtained by an oral enquiry from the heads of cultivating households. The data to be so obtained relate mainly to characteristics of farm population, livestock numbers, application of fertilisers and improved agricultural practices. Fortunately, for obtaining data regarding land utilisation, area under crops, irrigation, etc. it is not necessary to depend entirely up on the information given by the cultivators, a large number of whom are illiterate and do not maintain records The Technical Committee on ' Coordination of Agricultural Statistics" has recommended that the basic data on land utilisation and crops available in the *Khasra* Register could be re-tabulated holding-wise for giving this information. This method is possible over large areas except Kerala, West Bengal and Orissa where detailed land records are not available. In these States, the Census could provide the necessary bench-mark data which could be useful in introducing the much-needed improvements in the sample surveys.

ON THE CONSTRUCTION AND ANALYSIS OF A CLASS OF BALANCED ASYMMETRICAL FACTORIAL DESIGNS

By

K. KISHEN* AND B.N. TYAGI*

Summary. In experimentation in the biological, physical, chemical and other sciences, when a number of factors have to be tested simultaneously in a factorial scheme, situations arise when the numbers of levels of the factors to be tested are not all equal. Asymmetrical factorial designs in which the numbers of levels of the various factors are not all equal, are, therefore, of importance to experimenters in the various sciences.

This paper presents a method of construction, by the use of finite geometries, of the confounded balanced asymmetrical factorial (BAF) design $s^2 \times t$, where s is a prime positive integer or a power of a prime and t any positive integer less than s, and also the method of analysis of this design by fitting constants by the method of least squares. The design and analysis of $5^2 \times 2$, $5^2 \times 3$, $5^2 \times 4$, $7^2 \times 2$, $7^2 \times 3$, $7^2 \times 4$, $7^2 \times 5$ and $7^2 \times 6$ designs, which are particular cases of this general design, have been presented.

1. Introduction. The method of finite geometries first developed by Bose and Kishen (1940) for solving the problem of confounding in the general symmetrical factorial design s^m, where s is a prime positive integer or a power of a prime and m any positive integer, was extended by Kishen and Srivastava (1959*a*, 1959*b*) to the construction of confounded balanced asymmetrical factorial (BAF) designs of the type $s_1 \times s_2 \times \ldots\ldots \times s_m$, where s_1 is a prime positive integer or a power of a prime, m any positive integer, s_i's ($i=1, 2, 3, \ldots\ldots m$) are not all equal and $s_i \leqslant s_1$ for $i=2, 3, \ldots \ldots m$. This has been done by using curvilinear spaces or hypersurfaces in the m-dimensional finite Euclidean geometry $EG\,(m, s)$ constructed from the Galois Field $GF\,(s)$ and truncating the $EG\,(m, s)$ suitably.

The method of constructing BAF designs by the use of finite geometries has been briefly discussed. An important class of asymmetrical factorial designs $s^2 \times t$, where s is a prime positive integer or a power of a prime and t any positive integer less than s, has been considered and its method of analysis worked out in detail. As special cases of this general design, the design and analysis of $5^2 \times 2$, $5^2 \times 3$, $5^2 \times 4$, $7^2 \times 2$, $7^2 \times 3$, $7^2 \times 4$, $7^2 \times 5$ and $7^2 \times 6$, which are of importance in experimentation, have been given in detail.

2. Balanced Asymmetrical Factorial (BAF) designs. A design is said to be balanced if all the elementary treatment contrasts are estimated with equal precision. In confounded factorial design, however, we require more precision on main effects and lower order interactions than on higher order interactions. Kishen and Tyagi (1964) have termed a confounded factorial design balanced if the loss of infor-

---

*Department of Agriculture, U.P. Lucknow.

*Formerly Department of Agriculture, U.P., Lucknow, and presently I.A.R.S. New Delhi-12.

mation on each of the single degrees of freedom (d.f.) belonging to a partially confounded main effect or interaction is equal. Thus, if in a factorial design, effects F_1 and F_2 carrying n_1 and n_2 d.f. respectively are partially confounded, the design would be called balanced if losses of information on each of the n_1 d.f. belonging to F_1 are all equal, say, to d_1 and those on n_2 d.f. belong to F_2 to d_2 ; d_1 and d_2 need not be equal. Throughout this paper, we have called a confounded asymmetrical factorial design balanced in this sense.

3. Geometrical Method of Constructing BAF Designs.

Let $\alpha_0, \alpha_1, \ldots\ldots, \alpha_{s\cdot 1}$ denote the s elements of the Galois Field GF(s), where $s = p^n$, p being a prime positive integer and n any positive integer. Then the equation

$$f_1(x_1) + \alpha_{\mu_2} f_2(x_2) + \alpha_{\mu_3} f_3(x_3) + \ldots\ldots + \alpha_{\mu m} f_m(x_m) = \alpha_r \qquad \ldots(3.1)$$

represents a hypersurface in $EG(m, s)$, α_{μ_2} , α_{μ_3} ,.........$\alpha_{\mu m}$, α_r being any elements of $GF(s)$ and

$$f_1(x) = \alpha_{i1x} + \alpha_{i2} x^2 + \ldots\ldots \alpha_{i,s-1} x^{s-1} \qquad \ldots(3.2)$$

where α_{ij}, s ($i = {}^1$, 2, 3,......, m ; $j = 1$, 2,... ..$s-1$) are also elements of $GF(s)$. As r varies from 0 to $s-1$ in (3.1) we get a pencil of s hypersurfaces similar to the pencils of linear flats when

$$f_i(x) = x (i = 1, 2, 3, \ldots\ldots, m) \qquad \ldots(3.3)$$

It has been demonstrated by Kishen and Srivastava (1959b) that it is always possible to obtain an appropriate polynomial $f_i(x)$ such that for $x = \alpha_0, \alpha_1, \ldots\ldots \alpha_{s-1}$, $f_i(x)$ assumes only k distinct values.

Consider now m factors A_1, A_2, ...A_m at levels s_1, s_2......s_m respectively, where s_1 is a prime positive integer or a power of a prime and $s_i \leqslant s_1$ for $i = 2, 3, \ldots\ldots, m$ the s_i's($i = 1, 2, \ldots\ldots, m$) being not all equal. As it is always possible to have $s_i \leqslant s_1$ distinct values by taking a polynomial $f_i(x)$ with elements in $GF(s_1)$, let the factors A_1, A_2,........, A_m correspond to $x_1, f_2(x_2), \ldots\ldots\ldots, f_m(x_m)$ respectively. Suppose we now desire to confound paritally an m-factor interaction, we then take the pencil of hypersurfaces represented by

$$x_1 + \alpha_{\mu_2} f_2(x_2) + \alpha_{\mu_3} f_3(x_3) + \ldots\ldots + \alpha_{\mu m} f_m(x) = \alpha_r \qquad \ldots(3{\cdot}4)$$

where $u_2, u_3, \ldots\ldots u_m$ are fixed and $r = 0, 1, 2 \ldots\ldots s_1 - 1$, in the truncated $EG(m, s_1)$, it being assumed that x_1 varies over $\alpha_0, \alpha_1, \ldots\ldots \alpha_{s-1}$ and $x_i (i = 2, 3, \ldots\ldots, m)$ varies over the s_i elements, $\alpha_0, \alpha_1, \alpha_2, \ldots\ldots, \alpha_{u_{S_{i-2}}}$ of $GF(s_1)$ which provide the s_i distinct values for the polynomial $f_i(x_i)$. Each of the s_1 hypersurfaces in (3.4) will contain $s_2 \times s_3 \times \ldots\ldots \times s_m$ points to which would correspond $s_2 \times s_3 \times \ldots\ldots \times s_m$ treatment combinations. The pencil of hypersurfaces in (3.4) would thus divide the treatment combinations symmetrically in s_1 sets of $s_2 \times s_3 \times \ldots\ldots \times s_m$ treatment combinations each.

It has been shown by Kishen and Srivastava (1959b) that in the replication provided by the pencil of hypersurfaces (3.4), the interactions partially confounded are of two types, *viz.*, (a) the m-factor interaction corresponding to (3.4) which has been deliberately partially confounded, and (b) other interactions which would be partially confounded automatically owing to the fact that the number of combinations of levels of factors to which they relate is not equal to, or a factor of, the block size. Thus, in the replication corresponding to (3.4), the main effects of A_1 and all the interactions in which it enters will be partially confounded if

$$s_1 > s_i \ (i=2, 3, \ldots, m)$$

As has been shown by Kishen and Srivastava (1959*b*), to which the interested reader is referred for full details, the completely balanced asymmetrical factorial design would be obtained by varying $u_2, u_3, \ldots u_m$ over $1, 2, \ldots, s_1-1$ in (3.4). This would generate $(s_1-1)^{m-1}$ pencils of hypersurfaces and complete balance would be achieved in $(s_1-1)^{m-1}$ replications.

The above results have been utilized in constructing the *BAF* designs $s^2 \times t$, s being a prime positive integer or a power of a prime and t any positive integer less than s, discussed in the subsequent Sections.

4. Analysis of BAF designs $s^2 \times t$ $(t<s)$

The *BAF* design $s^2 \times t$ $(t<s)$, where s is a prime positive integer or a power of a prime, is given by the pencil

$$x_1 + x_2 + a_u f(x_3) = \alpha_r (r=0, 1, \ldots, s-1, u=1, 2, \ldots, s-1)$$

where $f(x_3)$ is varied from 0 to $s-1$, $f(x_3)$ assumes only t distinct values. Let us denote by x_i those treatment combinations of the factors A and B which correspond to the s points satisfying the equation

$$x_1 + x_2 = \alpha_i (i=0, 1, \ldots, s-1)$$

We shall analyse this design by fitting constants by the method of least squares. The sum of squares for the effects which are not confounded would be computed as usual. The partially confounded interactions are only *AB* and *ABC*. Let the constants to be fitted be chosen according to the following scheme :

General mean : μ

Blocks : $b_{11}, b_{12}, \ldots, b_{1s} ; b_{21}, b_{22}, \ldots, b_{2s} \ldots b_{s-1, 1} ; b_{s-1, 2}, \ldots b_{s-1, s}$

with $$\sum_{j=0}^{s} b_{ij} = 0 (i=1, 2, \ldots s-1)$$

Interaction *AB* : $(x_0), (x_1), \ldots, (x_{s-1})$ with $\sum_{i=0}^{s-1} (x_i) = 0$

Interaction *ABC* : $(x_0 c_0), (x_0 c_1) \ldots, (x_0 c_{t-1}), \ldots, (x_{s-1}, x_0)$

$(x_{s-1} c_1), \ldots (x_{s-1} c_{t-1})$ with

$$\sum_{i=0}^{s-1} (x_i c_j) = 0 (j=0, 1, \ldots t-1) \text{ and}$$

$$\sum_{j=0}^{t-1} (x_i c_j) = 0 (i=0, 1, \ldots, s-1)$$

Also, let $[X_i]$ denote the total for the interaction corresponding to

$(x_i)(i=0, 1, \ldots, s-1)$ and $[x_i c_j]$

the total for the interaction corresponding to

$$(x_i c_j)(i=0, 1, \ldots\ldots, s-1 ; j=0, 1, \ldots t-1.)$$

The normal equations for determining the constants $(x_i c_j)$ and (x_i) then come out as under :

$$tR_{ij}=s(st-s-t)(x_i c_j)+s^2(t-1)(x_i)+st(s-1)(c_j)$$
$$(i=0,\ 1,\ldots,s-1,\ \ j=0,\ 1,\ldots,\ t-1)$$
$$tQ_i=ts^2(t-1)(x_i)(i=0,\ 1,\ldots,\ s-1) \qquad \ldots(4.1)$$

where

$tR_{ij}=t[x_i C_j]-$[Totals of blocks containing treatments combinations $x_i c_j$]

and

$$tQ_i=\sum_{j=0}^{t-1} tR_{ij}$$

Then the estimate of $(x_i)(i=0,\ 1,\ldots,\ s-1)$ denoted by $(\hat{x_i})$, is given by

$$(\hat{x_i})=\frac{Q_i}{s^2(t-1)} \qquad \ldots(4.2)$$

From (4.2), it follows that the variance of an orthogonal contrast carrying one d.f. belonging to the interaction AB is $\sigma^2/s^2(t-1)$. In an unconfounded design, the corresponding variance would be $\sigma^2/st(s-1)$. Hence the loss of information on each of the $(s-1)$ partially confounded d.f. belonging to AB is given by

$$L[AB]=\frac{s-t}{t(s-1)}$$

The sum of squares for the $(s-1)$ partially confounded d.f. belonging to AB is

$$\frac{1}{s^2t^2(t-1)}\sum_{i=1}^{s-1}(tQ_i)^2 \qquad \ldots(4.3)$$

For computing the sum of squares corresponding to $(s-1)(t-1)$ partially confounded d.f. belonging to ABC, we calculate the quantities $T_i^k(i=0,\ 1,\ldots\ s-1)$; $k=1,\ 2,\ldots t-1)$ as under

$$T_i^k=\sum_{j=0}^{t-1} W_j k R_{ij} \qquad \ldots(4.4)$$

where $\sum_{j=0}^{t-1} W_j^k=0,$ and $\sum_{j=0}^{t-1} W_j k\ W_j k'=0$

for $k\neq k'=1,\ 2,\ldots\ t-1.$

From (4.1), it can he easily seen that the variance of an orthogonal contrast carrying one d.f. belonging to the interaction ABC is given by $tc^2/s(st-s-t)$. The corresponding variance in the unconfounded case would be $\sigma^2/s(s-1)$. Hence the loss of information on each of $(s-1)(t-1)$ partially confounded d.f. belonging to ABC is given by

$$L[ABC]=\frac{s}{(s-1)t}$$

The sum of squares corresponding to $(s-1)(t-1)$ partially confounded d.f. for ABC is

$$\frac{1}{st(st-s-t)}\sum_{k=1}^{t-1}\sum_{i=0}^{s-1}(tT_i^k-t\overline{T}^k)^2 \qquad ..(4.5)$$

where

$$s\bar{T}=k\sum_{i=0}^{s-1} T_i^k$$

The total loss of information on the (s−1) partially confounded d.f. of AB and $(s-1)(t-1)$ d.f. of ABC is

$$\frac{s-t}{t}+\frac{s(t-1)}{t}=s-1.$$

5. Analysis of BAF designs $s^2\times t$ ($s=5$, $t=2, 3, 4$; $s=7$, $t=2, 3, 4, 5, 6$).

Utilizing the results for the general BAF design $s^2\times t$ (s being any prime positive integer or a power of a prime and t any positive integer less than s) derived in the foregoing section, we have derived the analysis for some special cases of this design, *viz.*, $5^2\times 2$, $5^2\times 3$, $5^2\times 4$, $7^2\times 2$, $7^2\times 3$, $7^2\times 4$, $7^2\times 5$ and $7^2\times 6$ and present them in this Section.

5.1. Analysis of the BAF design $5^2\times 2$

The BAF design $5^2\times 2$ given by the pencil $x_1+x_2+\alpha x_3^4=c(\alpha=1, 2, 3, 4; c=0, 1,\ldots 4)$ is presented in Table 1.

TABLE 1

BAF designs $5^2\times 2$

Levels of C	*Replications*			
	I	*II*	*III*	*IV*
0	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$
1	$x_4x_0x_1x_2x_3$	$x_3x_4x_0x_1x_2$	$x_2x_3x_4x_0x_1$	$x_1x_2x_3x_4x_0$

Here $x_c(c=0, 1, 2, 3, 4)$ denotes the treatment combinations corresponding to $x_1+x_2=c(c=0, 1, 2, 3, 4)$.

Then, from (4.1), the normal equations for determining the constants (x_ic_j) and $(x_i)(i=0, 1,\ldots,4, ; j=0, 1)$ come out as under :

$$2R_{ij}=15(x_ic_j)+25(x_i)+40(c_j),$$

$$2Q_i=50(x_i), (i=0, 1,\ldots 4, j=0, 1) \qquad \ldots(5.2)$$

where $2R_{ij}=2\,[x_ic_j]-$[Totals of block totals containing treatment combination x_ic_j].

Then the estimate of (x_i), denoted by $(\hat{x}_i)(i=0, 1, 2, 3, 4)$ is given by

$$(\hat{x}_i)=\frac{Q_i}{25} \qquad \ldots(5.3)$$

From (5.3), the variance of an orthogonal contrast carrying one d.f. belonging to the interaction AB is $\frac{\sigma^2}{25}$ whereas in the unconfounded case this would be $\frac{\sigma^2}{40}$. Hence the loss of information on each of the 4 four partially confounded d.f. of AB is 3/8.

The sum of squares for the four partially confounded d.f. belonging to *AB* is

$$1/100 \sum_{i=0}^{4} (2Q_i)^2 \tag{5.4}$$

For computing the sum of squares for the four partially confounded d.f. belonging to *ABC*, we compute the quantities $T_i(i=0, 1, \ldots 4)$ as under

$$T_i=(R_{i0}-R_{i1}).(i=0,\ 1,\ldots 4) \tag{5.5}$$

From (5.2) it can easily be seen that the variance of an orthogonal contrast carrying one d.f. belonging to *ABC* is $2\sigma^2/15$. In the unconfounded case, however, the corresponding variance would be $\sigma^2/20$. Hence the loss of information on each of the 4 d.f. of *ABC* is 5/8. Thus, the total loss of information on the 4 d.f. of *AB* and the 4 d.f. of *ABC* is 4.

The sum of squares corresponding to the four partially confounded d.f. of *ABC* is

$$\frac{1}{30} \sum_{i=0} (2T_i-2\,\bar{T})^2 \tag{5.6}$$

where

$$5\bar{T}= \sum_{i=0}^{4} T_i$$

The analysis of the variance then takes the form as given in Table 2.

TABLE 2

Analysis of variance for BAF design $5^2 \times 2$

Source of variation	*D.F.*	*Sum of squares*
1. Replications	3	as usual
2. Block	16	,,
3. A	4	,,
4. B	4	,,
5. AB′	4	$1/100 \sum_{i=0}^{4} (2Q_i)^2$
6. AB²	4	as usual
7. AB³	4	,,
8. AB⁴	4	,,
9. C	1	,,
10. AC	4	,,
11. BC	4	,,
12. ABC′	4	$\frac{1}{30} \sum_{i=0}^{4} (2T_i-2\bar{T})^2$
13. AB²C	4	as usual
14. AB³C	4	,,
15. AB⁴C	4	,,
16. Error	131	by substraction

5.2. Analysis of the BAF design $5^2 \times 3$

The *BAF* design $5^2 \times 3$ given by the pencil

$$x_1+x_2+ax_3^2=c(a=1, 2, 3, 4\ ;\ c=0,\ 1, 2, 3, 4) \qquad (5.7)$$

is presented in Table 3.

TABLE 3

BAF design $5^2 \times 3$

Levels of C	Replications			
	I	II	III	IV
c_0	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$
c_1	$x_4x_0x_1x_2x_3$	$x_3x_4x_0x_1x_2$	$x_2x_3x_4x_0x_1$	$x_1x_2x_3x_4x_0$
c_2	$x_1x_2x_3x_4x_0$	$x_2x_3x_4x_0x_1$	$x_3x_4x_0x_1x_2$	$x_4x_0x_1x_2x_3$

From the (4.1), the normal equations for determining the constant $(x_i\ c_j)$ and (x_i) $(i=0, 1, 2, 3, 4\ ; j=0, 1, 2)$ come out as under :

$$3R_{ij}=35(x_ic_j)+50(x_i)+60(c_j)$$

$$3Q_i=150(x_i) \qquad i=0, 1,...4, j=0, 1, 2) \qquad ...5.8)$$

where

$$3R_{ij}=3[x_ic_j]-[\text{total of blocks containing treatment combination } x_ic_j]$$

and

$$3Q_i=\sum_{j=0}^{2} 3R_{ij}$$

Then the estimate of (x_i), denoted by $(\overset{\wedge}{x_i})$ is given by

$$(\overset{\wedge}{x_i})=\frac{Q_i}{50}(i=0,\ 1,\ 2,\ 3,\ 4) \qquad ...(5.9)$$

From (5.9), it follows that the variance of an orthogonal contrast carrying one d.f. belonging to the interaction *AB* is $\sigma^2/50$. In an unconfounded design, the corresponding variance would be $\sigma^2/60$. Hence the loss of information on each of the four partially confounded d.f. belonging to *AB* comes out to be 1/6.

The sum of squares for the four partially confounded d.f. belonging to *AB* is

$$\sum_{i=0}^{4} (3Q_i)^2 \qquad ...(5.10)$$

For working out the sum of squares for the partially confounded 8 d.f. belonging to *ABC*, let the quantities T_i' and T_i^1 $(i=0, 1,...4)$ be computed as follows ;

$$T_i^1=R_{i0}-R_{i1}$$

and

$$T_i^2=R_{i0}+R_{i1}-2R_{i2}.$$

Then the sum of squares for the partially confounded 8 d.f. belonging to *ABC* is

$$\frac{1}{105}\sum_{i=0}^{4}\{(3T_i^1-3\overline{T}^1)^2+(3T_i^2-3\overline{T}_i^{\ 2})^2\}$$

when $$5\overline{T}^k=\sum_{i=0}^{4} T_i^k \qquad \text{for } k=1,\ 2.$$

5.3. Analysis of the BAF Design $5^2\times 4$.

The *BAF* design $5^2\times 4$ given by the pencil

$$x_1+x_2+a(x_3+x_4^3)=c(a=1,\ 2,\ 3,\ 4\ ;\ c=0,\ 1,\ldots,\ 3) \qquad \ldots(5.11)$$

is presented in Table 4.

TABLE 4

BAF Design $5^2\times 4$

Levels of C	*Replications*			
	I	*II*	*III*	*IV*
c_0	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$	$x_0x_1x_2x_3x_4$
c_1	$x_3x_4x_0x_1x_2$	$x_1x_2x_3x_4x_0$	$x_4x_0x_1x_2x_3$	$x_2x_3x_4x_0x_1$
c_2	$x_2x_3x_4x_0x_1$	$x_4x_0x_1x_2x_3$	$x_1x_2x_3x_4x_0$	$x_3x_4x_0x_1x_2$
c_3	$x_1x_2x_8x_4x_0$	$x_2x_3x_4x_0x_1$	$x_3x_4x_0x_1x_2$	$x_4x_0x_1x_2x_3$

The normal equations for determining the constants (x_ic_j) and $(x_i)(i=0,\ 1,\ldots,4\ ;\ j=0,\ 1,\ 2,\ 3)$ come out as under :

$$4R_{ij}=55(x_ic_j)+75(x_i)+80(C_j),$$
$$4Q_i=300(x_i),$$

where

$$4R_{ij}=4[x_ic_j]-[\text{Total of blocks containing the treatment combination } x_ic_j]$$

and $$4Q_i=\sum_{j=0}^{3} 4R_{ij}$$

Then the estimate of (x_i) denoted by $(\hat{x_i})$ is given by

$$(\hat{x_i})=\frac{Q_i}{75}(i=0,\ 1,\ 2,\ 3,\ 4) \qquad \ldots(5.13)$$

As the variance of an orthogonal contrast carrying one d f. belonging to the interaction *AB* is $\sigma^2/75$ as compared to $\sigma^2/80$ in the unconfounded case, the loss of information on each of the four partially confounded d.f. of *AB* is 1/16.

The sum of squares for these four d.f. of *AB* is

$$\sum_{i=0}^{4} (4Q_i)^2/1200 \qquad \ldots(5.14)$$

For computing the sum of squares corresponding to the partially confounded 12 d.f. belonging to *ABC*, we calculate the quantities T_i^k ($i=0, 1, \ldots 4$, $k=1, 2, 3$) as under

$$T_i^1 = R_{i0} - R_{i1},$$

$$T_i^2 = R_{i0} + R_{i1} - 2R_{i2}$$

$$T_i^3 = R_{i0} + R_{i1} + R_{i2} - 3R_{i3},$$

and

$$5\overline{T}^k = \sum_{i=0}^{4} T_i^k \text{ for } k=1, 2, 3.$$

Then the sum of squares for the partially confounded 12 d.f. belonging to *ABC* is

$$\frac{1}{220} \sum_{i=0}^{4} \{(4T_i^1 - 4\overline{T}^1)^2 + (4T_i^2 - 4\overline{T}_i^2)^2 + (4T_i^3 - 4\overline{T}_0^3)^2\}$$

From (5.12), it is easily seen that the loss of information on each of the partially confounded 12 d.f. of *ABC* is 5/16, the total loss of information on *AB* and *ABC* being 4.

5.4. Analysis of the BAF design $7^2 \times t (t=2, 3, 4, 5, 6)$.

The pencils giving the *BAF* design $7^2 \times 2$, $7^2 \times 3$, $7^2 \times 4$, $7^2 \times 5$ and $7^2 \times 6$ are presented in Table 5.

TABLE 5

Pencils for the BAF design $7^2 \times t$ (t=2, 3, 4, 5, 6)

Sl. No.	*BAF design* 2	*Pencil* 3	
1.	$7^2 \times 2$	$x_1 + x_2 + a x_3^6 = c$	$a=1, 2, \ldots\ldots 6$; $c=0, 1, \ldots\ldots, 6$
2.	$7^2 \times 3$	$x_1 + x_2 + a x_3^3 = c$	,,
3.	$7^2 \times 4$	$x_1 + x_2 + a x_3^4 = c$	,,
4.	$7^2 \times 5$	$x_1 + x_2 + a(x_3^2 + x_3^3) = c$	,,
5.	$7^2 \times 6$	$x_1 + x_2 + a(x_3 + x_3^6) = c$	,,

In all the designs given in Table 5, the interaction *AB* and *ABC* are partially confounded. The analysis of variances follows from (4.1). The sum of squares for

SAMPLE SURVEYS ON FRUIT CROPS

By

G.R. SETH,
B.V. SUKHATME
AND
A.H. MANWANI

One of the basic requirements for proper planning and execution of programmes for increasing production of fruits is the availability of comprehensive and reliable statistics of area and yield of fruits and the information relating to the existing cultivation practices and status of fruit cultivation. No realistic targets can be fixed for the production of fruits, in the absence of reliable statistics of area and yield. These statistics are also required to assess the contribution of fruits towards the national income.

At present, the statistics of area for important categories of fruits such as mango, banana, citrus, grapes, and other fresh fruits are collected by various States through a reporting agency and are reported in their 'Season and Crop' reports. These statistics are consolidated and published by the Directorate of Economics and Statistics in "Estimates of Area and Production of Principal Crops in India". These are rough estimates and suffer from lack of standard definitions and concepts involved in the collection of area statistics. There is also considerable time lag in the availability of these statistics. So far as the yield statistics are concerned, practically no data are available for any part of the country. Thus, at present, the statistics of area and yield of fruits are incomplete and inadequate and lack the desired degree of accuracy and inter-State comparability.

The imperative need for collecting reliable statistics of area and yield of fruit crops has been emphasised from time to time since the establishment of the Fruit Development Board of the Indian Council of Agricultural Research and more recently by the Committee for improvement of Agricultural Statistics appointed by the Ministry of Food and Agriculture. The Committee recommended that in those States which are cadastrally surveyed and posses a primary reporting agency at the village level, the areas under different fruits of regional importance should be recorded separately along with the number of trees in the basic village form. Besides this, the Committee also recommended several other measures to ensure proper estimation of areas under individual fruits in the case of mixed orchards. These recommendations when implemented, are expected to improve considerably the statistics of area under fruit crops. But so long as these recommendations are not adopted the partially confounded 6 d.f. belonging to *AB* and $6(t-1)(t=2, 3, 4, 5, 6)$ d.f. belonging to *ABC* are presented in Table 6.

TABLE 6

Sums of squares for AB and ABC

Sl. No.	*Design*	*Estimate of* x_i	*S.S. for 6 d.f. of AB*	*S.S. for 6 (t—1) d.f. of ABCe*
1.	$7^2\times 2$	$1/49\ Q_i$	$1/196 \sum_{i=0}^{6} (2Q_i)^2$	$1/70 \sum_{i=0}^{6} (2T_i-2\bar{T})^2$
2.	$7^2\times 3$	$1/98\ Q_i$	$1/882 \sum_{i=0}^{6} (3Q_i)^2$	$1/231 \sum_{k=1}^{2} \sum_{i=0}^{6} (3T_i^k-3\bar{T}^k)^2$
3.	$7^2\times 4$	$1/147\ Q_i$	$1/2352 \sum_{i=0}^{6} (4Q_i)^2$	$1/476 \sum_{k=1}^{3} \sum_{i=0}^{6} (4T_i^k-4\bar{T}^k)^2$
4.	$7^2\times 5$	$1/196\ Q_i$	$1/4900 \sum_{i=0}^{6} (5Q_i)^2$	$1/805 \sum_{k=1}^{4} \sum_{i=0}^{6} (5T_i^k-5\bar{T}^k)^2$
5.	$7^2\times 6$	$1/245\ Q_i$	$1/8820 \sum_{i=0}^{6} (6Q_i)^2$	$1/1218 \sum_{k=1}^{5} \sum_{i=0}^{6} (6T_i^k-6\bar{T}^k$

REFERENCES

Bose, R.C, and Kishen, K, (1940). On the problem of confounding in the genera symmetrical factorial designs. Sankhya 5, 21-36.

Kishen, K and Srivastava, J.N. (1959). Confounding in asymmetrical factorial designs in relation to finite geometries. Current Science, **28**, 98-100.

Kishen, K and Srivastava, J.N. (1959*b*). Mathematical theory of confounding in the symmetrical and asymmetrical factorial designs. *Journal Ind. Soc. Agri. Stat. II.* 73-110.

Kishen, K and Tyagi, B.N. (1961). On some methods of constructions of asymmetrical factorial designs. Current Science 30. 407-409.

Kishen, K. and Tyagi, B.N. (1964). On the constructions and analysis of some balanced asymmetrical factorial designs. Cal. Stat. Assoc. Bull. 13, 123-149.

as a normal measure, it will be necessary to conduct sample surveys for estimating area statistics of fruit crops for which suitable sampling techniques will have to be developed. In those States which are not cadastrally surveyed and do not have any reporting agency, it will be necessary to conduct periodically a sample survey or a census to obtain estimates of area under fruit crops. In areas where fruits are grown extensively, it may be desirable to conduct sample surveys as recommended by the Committee. In those areas where fruit cultivation is not extensive, it will be necessary to conduct a census of fruit trees periodically. For determining yield rates and total production of cereal and other field crops, an objective method of crop-cutting experiments or random sample surveys has replaced the traditional subjective method of determining yield rates. As sampling techniques developed for crop estimation surveys on field crops will not be directly applicable to fruit crops, suitable sampling techniques will have to be developed for estimating yield rates and production of fruit crops which differ in several respects from field crops. Unlike field crops which are grown in regular fields, fruit trees are grown not only in orchards but on any suitable land, such as, on canal banks, field bunds, road side, backyard of houses, etc. Fruit trees take quite a few years before they start bearing fruits. The bearing age of a tree varies considerably from fruit to fruit. Again, all the trees in an orchard are generally not of the same age. In fact, it is quite common to find in the same orchard, trees belonging to different age groups as well as of different fruits. Also, all the trees of bearing age may not be bearing fruits during the season. The proportion of trees of bearing age failing to bear fruits, varies with age and differs for different fruits as well as from season to season. The harvesting of an orchard may extend over several weeks which is generally not the case with cereal crops. There are several fruits, such as citrus, guava, where there are two or more harvesting seasons in a year. All these features of fruit cultivation will, therefore, have to be carefully considered while evolving a sampling technique to estimate their extent of cultivation and production.

At the instance of the Ministry of Food and Agriculture, the Institute of Agriculture Research Statistics, I.C.A.R. initiated a series of sampling investigations on important fruit crops, such as mango, guava, banana, orange, lime, etc. with the following broad objectives in view:

(i) To evolve a sampling technique for estimating with desired degree of precision the area, yield rate and annual production of fruits, and

(*ii*) To collect reliable data on manurial and cultivation practices of the crop as practised by the orchardists.

The data collected during the course of these investigations included information on (*i*) area under the crop ; (*ii*) total number of orchards under bearing and non-bearing categories ; (*iii*) total number of trees under bearing and non-bearing categories as well as according to variety ; (*iv*) yield of sampled trees both in terms of weight and number of fruits during the entire harvesting season and (*v*) cultivation practices, such as spacing, source of material for planting, frequency of irrigation, manuring and - other cultural operations followed by the orchardists.

The Sampling design adopted for the Surveys

The sampling design adopted for fruit surveys may be described as two-phase multi-stage stratified random sampling with tehsils or groups of adjacnet tehsils as strata. Within each stratum, a specified number of villages roughly in proportion to area under fruits were selected at random according to a probability scheme of sampling. The probability scheme of sampling used in a survey depended on the auxiliary information available and differed from fruit to fruit. For example, in the case of lime survey in Andhra Pradesh, information was available in respect of area under lime for each village. In this survey, villages were therefore, selected with probability proportional to reported area under lime. In the case of surveys on mango and guava in Uttar Pradesh, information was available concerning area under kharif fruits for each village. Consequently, villages were selected with probability proportional to area under kharif fruits. In case, no auxiliary information was available or this information was not very reliable, the villages were selected with equal probability. All the selected villages were completely surveyed to obtain information concerning the extent of cultivation of the fruits under study as indicated by the number of orchards, area under fruit and number of trees according to variety under bearing and non-bearing categories. A sub-sample of the selected villages was randomly taken for the purpose of yield study. For this purpose a specified number of orchards (usually 3 to 5) growing the fruit under study were selected in each of the selected villages in the sub-sample. Within each selected orchard a specified number of bearing trees (usually 9 to 12) were selected at random in the form of clusters of 3-4 trees each to collect data on the yield of the fruit.

The data in respect of items relating to the extent of cultivation were collected by counting all the trees in the sampled villages. The data on yield of the sampled trees were collected by actual weighment of the produce during the entire harvesting season while, those on cultivation practices were collected by enquiry from the selected orchardists.

In this paper we shall confine our discussion with reference to the surveys carried out on mango and guava in Uttar Pradesh and lime in Andhra Pradesh. The survey on mango was carried out in three rounds from 1958-59 to 1960-61 of which the first two rounds were carried out in Saharanpur district, while, the third round was carried out in Varanasi district of Uttar Pradesh. The survey on guava was carried out in three rounds from 1959-60 to 1961-62 in Allahabad district of Uttar Pradesh. The survey on lime was carried out in three rounds from 1961-62 to 1963-64 in Nellore district of Andhra Pradesh. The number of villages, the orchards and the trees selected for the purpose of estimating the acreage, yield rate and production in these surveys are given in Table 1.

TABLE 1

Fruit	*District*	*Year*	*No. of villages selected for*		*No. of orchards selected for yield and cultivation practices*	*No. of trees selected for yield*	*Method of selection of primary units in each stratum*
			enumeration	*yield*			
1	2	3	4	5	6	7	8
Mango	Saharanpur	1959-60	50	50	232	2,556	P.P.S.
	Varanasi	1960-61	140	56	307	2,267	P.P.S.
Guava	Allahabad	1959-60	51	51	255	3,060	P.P.S.
		1960-61	148	58	277	3,321	P.P.S.
		1961-62	191	57	285	3,406	S.R.S.
Lime	Nellore	1961-62	100	40	226	1,800	P.P.S.
		1962-63	124	48	202	1,536	P.P.S.
		1963-64	70	39	253	1,896	S.R.S.

Results of the Sample Surveys

Table 2 summarises the main results obtained from the sample surveys on mango, guava and lime.

It will be seen from the table that the estimates of area and number of trees were obtained with standard error ranging between 6 to 10 per cent in mango and guava surveys while, in the case of lime survey the standard error was of the order of 3 per cent only. The main factor contributing towards substantial reduction of the percentage standard error in lime survey was the selection of primary units with probability proportional to area under lime. Although, varying probability selection was adopted in mango and guava surveys, the character for probability selection in these surveys was area under all the kharif fruits which was not highly correlated with the area Under the fruit under study. The detailed results of investigations on the data collected from the pilot surveys are discussed below.

TABLE 2

Estimates of different characters in the surveys on mango, guava and lime

Characters	*Fruit Crop*							
	Mango in Saharanpur		*Mango in Varanasi*		*Guava in Allahabad*		*Lime in Nellore*	
	Est.	*% S.E.*	*Est.*	*% Est.*	*Est.*	*% S.E.*	*Est.*	*% S.E.*
1. Gross area* under crop (in hectares)	9,414	9·7	9,582	7·9	2,262	7·9	1,689	3·2
2. Total no. of trees (in thousands)	612	10·1	392	7·4	578	6·7	489	3·4
3. Average yld. per tree of bearing age (in kg)	42·2	8·4	40·4	9·0	17·6	5·3	69·4	3·8
4. Av. yield per bearing tree (in kg)	53·1	7·2	45·9	7·8	16·6	5·1	69·9	3·8
5. Totol production (in thousands tonnes)	16·7	12·0	11·6	13·0	8·0	6·8	17·5	4·8

Efficiency of Stratification

Increasing efficiency of estimates through stratification requires that the units within strata should be as similar as possible with regard to the characters under study. As geographically contiguous areas are likely to have similar cultivation practices as well as similar climate and soils, geographical contiguity suggests itself as a suitable criterion for stratification for the study of area and yield of fruit crops. Such stratification can also be adjusted to meet the needs of obtaining estimates for smaller divisions like districts or groups of taluks required for planning of horticultural development programmes. Besides, availability of frame, administrative convenience and organization of field work dictate the constitution of geographically contiguous areas into strata. In the pilot studies discussed in this paper tehsils within the district under survey were taken to be strata. The results indicated that formation of tehsils as strata increased the precision of the estimates of number of trees by 133 per cent for guava crops and by 235 per cent for the mango crop. However, corresponding gains in precision of the

---

*Gross area under the crop was defined as the area of all the orchards in which the given fruit crop was grown. This also included the area under other crops grown in the orchards along with the given fruit crop.

estimates of average yield per bearing tree for both these surveys was rather small. Similar results in gain in efficiency were also observed for lime crop.

As regards optimum allocation of units to different strata, it depends upon coefficient of variation of the primary units within strata. Coefficient of variation of villages in respect of number of trees varied little from stratum to stratum except in the case of guava survey where it varied from 69 per cent to 116 per cent. This suggests that optimum allocation of primary units (villages) among strata may be made in proportion to the sizes of the strata as determined by the number of trees. As, the information regarding the number of trees is not generally available, allocation among strata might be done in proportion to the size of some of the character highly correlated with the number of trees, such as latest available reported area under the fruit crop under study or the area under all the fruit crops. Such an allocation in proportion to reported area under fresh fruits was attempted in these investigations and gave estimates of number of trees which were 94 per cent and 85 per cent efficient as compared to the estimates based on Neyman optimum allocation for mango and guava respectively.

Probability Selection of Units

The next important problem is to decide on probability scheme of sampling to select primary units viz., villages within a stratum. It is well known that considerable gains in the precision of estimates is achieved by selecting villages with probability proportional to the value of some character which is highly correlated with the yield of the crop under study. The characters which suggest themselves are the area under the fruit in question, total number of trees, area under fruits, etc. In surveys on mango and guava, information was available concerning the area under kharif fruits. In these surveys, villages were, therefore, selected with probability proportional to area under kharif fruits. The following table gives the gain due to sampling with varying probability over simple random sampling with respect to the character (1) total number of trees and (2) average yield per tree in the case of the survey on mango carried out in Varanasi district of U.P.

TABLE 3

Percentage gain in efficiency due to sampling of villages with varying probabilities over simple random sampling with stratification as adopted in the survey.
(Varanasi District—Mango)

	Character under study	
Tehsil	*No. of trees*	*Av. yield per tree*
Varanasi	150	106
Gyanpur	150	41
Chandoli	669	185
District	169	130

It will be observed that selecting villages with probability proportional to area under kharif fruit has resulted in substantial gain in precision in respect of both the characters.

In the case of Allahabad survey on guava, the villages in the first two rounds were selected with replacement with probability proportional to area under kharif fruits while, they were selected with equal probability and without replacement in the last round of the survey. The Table 4 gives percentage standard errors for the estimates of gross area, number of guava orchards, number of guava trees, net area under guava and average yield per tree obtained from the last two rounds of the survey.

TABLE 4

Percentage standard errors of different estimates when the villages are selected with varying probabilities and with equal probabilities

Character	% S.E.	
	Varying probabilities *1960-61*	*Equal probabilities* *1961-62*
Gross area under guava	6·2	8·6
No. of guava orchards	5·6	8·7
No. of guava trees	5·2	9·0
Net area under guava*	6·5	9·5
Average yield per tree	7·5	10·9

It will be seen that the percentage standard errors in the case of sampling with varying probabilities is always less as compared to that from sampling with equal probability and without replacement even though, the size of the sample in 1961-62 was somewhat larger as compared to 1960-61. The results suggest that it is preferable to select units with probabilities proportional to area under kharif fruits instead of selecting them with equal probability and without replacement.

---

* Net area under the fruit was defined as area under the given fruit [crop only excluding area under trees of other fruit crops planted in the orchards. This area was estimated on the basis of average spacing between adjacent trees.

Determination of sample size

It was observed that there was considerable variation between villages in regard to area under the crop and the number of bearing trees. This suggests that quite a large number of villages will have to be taken for estimating these characters. Table 5 below gives the number of villages required to be sampled to estimate the number of bearing trees with different degrees of precision. It will be seen that about 300 villages will be required to obtain an estimate of the total number of bearing trees with a fair degree of precision.

TABLE 5

No. of villages required for estimating the number of bearing trees with given precision

Per cent Standard error	*Mango Survey (selection with probabilities proportional to reported area under orchards)*	*Guava Survey (selection with probabilities proportional to reported area under kharif fruts)*	*Lime Survey (selection with probabilities proportional to reported area under lime)*
5	328	335	253
6	228	233	176
7	167	171	130

So far as the estimate of average yield per tree is concerned the contribution to its standard error arises from three different sources : (*a*) variation between villages, (*b*) variation between orchards within villages, and (*c*) variation between trees within orchards. Of these, the variation between villages was found to be the most important component of the total variation. The number of orchards per village and the number of trees per orchard required to estimate the average yield per tree depend on the variation between orchards and the variation between trees within orchards. It was found that for the purpose of estimating the average yield per tree, it is enough to select 3 to 5 orchards per village. Increasing the number of orchards beyond five does not seem to decrease appreciably the standard error of the estimate. Similarly, it has been found that it is sufficient to select 9 to 12 trees per orchard. Table 6 gives number of villages required to be sampled for estimating the average yield per tree with five per cent standard error for varying number of orchards to be selected from each village and number of clusters of trees to be selected from each selected orchard.

On the basis of these results, it will be seen that a sample of about 130 villages with five orchards per village and 12 trees per orchard will be necessary to estimate the average yield per tree with five per cent standard error. However, if the villages are selected with probability proportional to area under the fruit crop under study, as was done in the case of lime survey, a much smaller number of villages need be taken to estimate the average yield per tree.

TABLE 6

Estimates of number of villages required to estimate average yield per tree with five per cent standard error as obtained in different surveys for varying number of orchards and number of clusters of trees

	Mango Survey (selection of villages with probabilities proportional to area under orchards)				*Guava Survey (selection of villages with probabilities proportional to area under kharif fruits)*				*Lime Survey (selection of villages with probabilities proportional to area under lime)*			
*Orchards/ Clusters**	*3*	*4*	*5*	*6*	*3*	*4*	*5*	*6*	*3*	*4*	*5*	*6*
1	158	133	119	109	200	178	165	155	99	72	61	54
2	155	131	117	108	180	163	153	146	50	46	40	36
3	154	130	116	107	173	158	149	144	45	38	33	31

Double Sampling

As noted earlier, the number of villages required to estimate the average yield per tree is considerably smaller than that required for estimating the total number of bearing trees. As both the average yield per tree and the total number of bearing trees have to be estimated with sufficient degree of precision to obtain reliable estimate of the total production, it follows that a larger sample of villages will have to be selected for estimating the total number of bearing trees, while a sub-sample of these villages may be selected for estimating the average yield per tree. The method of double sampling can therefore be used with advantage in fruit surveys. This method has been found to be statistically efficient and has resulted in gain in efficiency upto 80 per cent or even more for estimating the total production of fruits.

Even though, the pilot surveys have demonstrated the feasibility of obtaining reliable estimates of area and yield of fruit crops through sample

---

* Cluster size is assumed to be three in the case of mango and four in the case of guava and lime.

ROLE OF STATISTICS IN STANDARDIZATION

by

B.N. SINGH*

Statistics has an important role to play in standardization. When properly used, it provides sound, satisfactory and economic solutions to a variety of problems encountered in the formulation and implementation of standards. This paper illustrates some applications of the statistical concepts and techniques to various aspects of standardization with a special reference to the experience gained within the framework of ISI.

Since statistics deals essentially with the collection, analysis and interpretation of the data obtained as a result of the interplay of multiplicity of factors, it would be applicable to all such phases of standardization where such data are likely to arise. Thus, they would be immensely useful in the specification of quality, including the limitation of variety, in prescribing test methods for the evaluation of quality and in formulating sampling plans for appraising conformance to the specified quality.

Besides, in the course of preparation of standards, it may sometimes become necessary to conduct investigations with a view to collecting relevant information. Statistics is also helpful in designing such investigations and interpreting the data resulting therefrom.

The importance of statistical approach in standardization work, it may be mentioned, has also been realized at the international level. The Council of the International Organization for Standardization (ISO), Standing Committee for the Study of Scientific Principles of Standardization (STACO) and the Commonwealth Standards Conference have already made recommendations advocating greater use of statistical methods in various aspects of standardization.

Specification of Quality

While prescribing the specification requirements for a material, the ever-present question is how to secure agreement among diverse interests of the quality levels for the material. The statistical analysis of the quality data, collected in a systematic manner from different manufacturing units, has been found to help considerably in securing this agreement. Such data are often available with manufacturers and with little extra effort they could be summarized for the guidance of technical committees. The revision of the Indian

---

*Indian Standards Institution, New Delhi.

survey technique, the conduct of State-wide surveys to provide reliable statistics of all the major fruit crops grown in a state will give rise to new problems both technical as well as organisational. It will be necessary to tackle these problems before the technique can be recommended to the State for adoption on a routine basis. As a first step in this direction, the Institute of Agricultural Research Statistics has undertaken one State-wide survey in Andhra Pradesh and another survey on temperate fruits in Kumaon region of Uttar Pradesh and Himachal Pradesh. The analysis of the data collected from these surveys is in progress. It is hoped that the results obtained from these surveys will help in finalising the planning, designing and conduct of such surveys and further confirm the feasibility of employing sample survey technique for collection of reliable statistics on fruit crops.

Standard specification for DDT Dusting Powders (IS: 564-1961) may be cited as a typical example. This standard had been originally issued in 1955 with a tolerance of ± 2.5 per cent on the declared value of the DDT content, which at that stage was generally agreed to be reasonable and practical of achievement. Later, however, several manufacturers complained that they were not in a position to meet this tolerance of ±2.5 per cent even after introducing adequate control during production and desired it to be reviewed. Statistical analysis of the production data of three leading manufacturers of the country showed that their coefficient of variation for DDT technical content was of the order of 1·6, 1·8 and 2·6 per cent respectively. A tolerance of ±2·5 per cent allowed a coefficient of variation of 1·25 per cent only, indicating that none of the manufacturers could meet the specified limits for DDT content. Since the coefficient of variation of 2.5 per cent covered a major part of production and since it called for the tolerance of ±5 per cent, the Committee accordingly agreed to modify the tolerance for DDT technical content from ±2.5 per cent to ±5 per cent of the declared value in the revised version of the specification.

For personal safety against electric shock, the permissible limit of the leakage current flowing through the exposed parts of domestic electrical appliance has to be specified. But the sensation of shock depends, among other factors, upon environmental conditions, including climatic factors and personal habits of user groups. The permissible limits of current are, therefore, expected to vary for different regions. Statistically designed experiments are perhaps the only means for arriving at the satisfactory limits in such circumstances. The IEC Publication 'Safety Requirements for Mains-operated Radio Receiving Apparatus (Pub 65:1952)', had specified the maximum permissible current through a live part as 700 microamperes peak. This value was considered to be too high for tropical countries like India, where, apart from high temperature and humidity, people often go bare-footed on uncovered floors. In order to arrive at the limits acceptable under such conditions, a statistically designed experiment was conducted throughout India covering various temperatures and humidity conditions. The experiment consisted in subjecting groups of people in different regions to different levels of current in a random order and recording their reactions as unpleasant or otherwise. The results showed that 50 per cent of the subjects reported unpleasant sensation of shock at 700 microamperes. At the current level of 300 microamperes, only 2 per cent of the subjects complained about sensation of shock, which led the relevant IEC committee to recommend 300 microamperes as the safe value for permissible leakage current under tropical conditions.

One of the objectives of standardization is simplification or limitation of varieties of products. This objective can be economically and efficiently achieved if the data are collected on the demand for each type and size of the product in terms of the quantity purchased. Any analysis of such data, with a view to evolving the frequencies for the various demands, would show the types or sizes which can substantially meet the demand in the market. Other sizes, which have very little demand, can be discarded, thereby resulting in great economy in the cost of production, storage and handling. Such an analysis was recently carried out by ISI in connection with rationalization of overall dimensions of motor vehicle batteries. Two specifications dealing with motor vehicle batteries, namely, IS:395-1959 Lead-acid storage batteries (light duty) for motor vehicles (revised), and IS:985-1958 Lead-acid storage batteries (heavy duty) for motor vehicles, had specified only the maximum dimensions in the former case and no dimensions in the latter case. This resulted in as many as 232 different sizes of batteries, thereby necessitating the use of a large number of moulds for their manufacture. An analysis of the data on the overall dimensions collected from various battery manufacturers and on the need for different sizes required by the automobile manufacturers revealed the possibility of reducing the number of sizes. A proposal was, therefore, made to the relevant committee to reduce the number of sizes. The Committee, after considering the data, decided that only 11 sizes for all types of batteries could meet most of the demand for the new automobiles and included the same in the latest revision of the specification. The problem of sizes for replacement purposes was not considered as such sizes are expected to become obsolete in course of time.

In specifying dimensional tolerance on an assembly consisting of two or more parts individually made to specify tolerances, the engineering practice has been to arrive at the assembly limits by assuming that maximum deviations of the parts will determine the maximum deviation of the assembly in 100 per cent of the cases. Similar approach is also followed for the combination of two or more quantities with their individual errors. Thus, the tolerance (or the error) of the following assemblies (or combinations) of two parts $A \pm T_1$ and $B \pm T_2$ are generally evaluated as:

	Assembly	*Tolerance*
(*a*)	$A \pm B$	$\pm (T_1 + T_2)$
(*b*)	$A \times B$	$\pm (BT_1 + AT_2)$
(*c*)	A/B	$\pm \frac{1}{B} (BT_1 + AT_2)$

The application of the statistical principles in the manufacture of the parts or the operation of the errors shows that, within the tolerance or error limits, the frequency distribution of the units produced follow the Normal (Gaussian) Law, according to which approximately 68 per cent of the parts lie within $\pm 1/3$ (tolerance or error), 95 per cent within $\pm 2/3$ (tolerance or error) and so on. An assembly (or a combination) of two parts with maximum tolerance (or error) is, therefore, rare and the extreme limits of tolerances (or errors) given by the above formulae are not of much practical significance. For assemblies (or combinations) referred to above, the tolerances (or errors) obtained from the following formulae will cover practically all the assemblies (or combinations)

Assembly	*Statistical Tolerance*
(*a*) $A \pm B$	$\pm \sqrt{T_1^2 + T_2^2}$
(*b*) $A \times B$	$\pm \sqrt{B^2T_1^2 + A^2T_2^2}$
(*c*) A/B	$\pm \frac{1}{B^2} \sqrt{B^2T_1^2 + A^2T_2^2}$

It will be seen that the tolerance on an assembly derived from statistical principles is much narrower than that obtained from the usual engineering practice. These considerations, therefore, provide the necessary scope for relaxing tolerances on individual parts or assembly or conversely obtaining from accurate assemblies from parts with existing tolerances, both leading to greater economy in manufacture.

Statistical tolerances on assembly $A \pm B$ have been well discussed in the literature on the subject. In the efforts to propagate these concepts among ISI Committees, need was felt for appropriate formulae for assembly or combination $A \times B$ and A/B, which had to be specially developed.

Evaluation of quality

Method for measurement (or evaluation) of quality is another important aspect of standardization wherein statistical methods would be extremely useful. When repeated determinations are made by a test method, the observations are found to vary. It is important that the variation between these duplicate determinations be known in quantitative terms so as to serve as a guide to the users of the test methods for judging the acceptability of the test results. The magnitude of this variation when determinations are made in the same or different laboratories are of practical significance. Numerical estimates for these variations are expressed in terms of repeatability and reproducibility which may be defined as follows :

> *Repeatability* is a quantitative measure of the variability associated with a single operator in a given laboratory obtaining successive repeat results on the same apparatus. It is defined as the difference between two such single results that would only be exceeded in the long run in one case in twenty in the normal and correct operation of the test method.
>
> *Reproducibility* is a quantitative measure of the variability associated with operators working in different laboratories, each obtaining single result on identical test materials. It is defined as the difference between two such single and independent test results that would be

exceeded in the long run in only one case in twenty in the normal and correct operation of the test method.

To determine these values for any test method, statistically designed experiments have to be conducted by different operators in different laboratories so that valid conclusions could be drawn. As a result of this type of investigations. it has been possible to specify in IS : 1351-1959 Methods of test for coal and coke, values of repeatability and reproducibility for the methods of determination of ash and total sulphur. In IS : 1448 (Part I)-1960 Methods of test for petroleum and its products, these values have been adopted from the IP and ASTM standards. However, investigations have now been undertaken for the determination of the repeatability and reproducibility values on the basis of inter-laboratory collaboration under Indian conditions.

The knowledge of repeatability and reproducibility is also helpful in the comparison of two or more test methods for selection. For, if a number of test methods are available for the same purpose, the one with the best repeatability and reproducibility values could be chosen for prescribing in the standard. Alternatively, if the test methods have more or less the same repeatability and reproducibility, the one which is more economical and/or less time consuming could be preferred.

With a view to propagating the concepts of repeatability and reproducibility on a wider basis to various industries and its technical committees, ISI has undertaken the formulation of an Indian Standard on method for the determination of the precision of the test methods.

Inspection of Quality

The inspection of quality with a view to judging the conformity of a lot of manufactured items to the relevant specification, is an important aspect which needs attention in the formulation of standards. For example, in the case of bicycle hubs it would be necessary to test them for their finish, dimensions, concentricity, free running and correct assembly. In many cases, like those in which destructive testing has to be done, it is not possible to test each and every product in the lot, and hence inspection has necessarily to be carried out on sampling basis. While selecting a sample it is necessary to ensure that its size is neither too small to afford very little protection to the consumer nor too large to make the testing uneconomical. The right balance between these two types of interests can be achieved only by evolving sampling plans on the basis of statistical considerations.

Though great advances have been made during the past few years in the theory of sampling, their applications to practical problems still present difficulties because of the varied nature of materials, trade practice and economic considerations. These difficulties multiply further when sampling plans are required to be specified in standards which are intended to cater to the needs of very wide groups of people with varying interests. However attempts are being constantly made by ISI to include suitable sampling plans in Indian Standards dealing with material specifications and test methods. In order that these sampling plans may be easily adopted by trade and industry, care is taken to evolve them by suitably blending the theoretical requirements with practical considerations.

Since production data on control charts always provide better statistical evidence regarding conformity of the material to the specification requirements than an acceptance sampling plan, the Indian Standards, wherever appropriate, recommended to the manufacturer to make such data available to the consumers for appraising the conformity of their purchases. Mention may be made in this connection of IS : 1528-1962 Methods of sampling and physical tests for refractory materials, IS : 3704-1966 Methods for sampling of light metal and their alloy products, etc.

To assist the manufacturers in quality control during production, ISI has published an Indian Standard method for statistical quality control during production by the use of control chart (IS : 397-1952). This standard is, however, under revision with a view to incorporating some of the recently developed control charts for median, mid-range and the largest and smallest observations which are easier to operate and at the same time quite efficient for small samples.

To help ISI committees in overcoming the sampling difficulties, as also to acquaint them with the principles of sampling, ISI has issued IS : 1548-1960 Manual on basic principles of lot sampling. To meet a wide variety of requirements for sampling inspection plans, ISI has published an Indian Standard on sampling inspection Tables (IS : 2500). Part I of this standard deals with inspection by attributes and by count of defects, whereas Part II deals with inspection by variables.

According to the attributes plans, an item is to be classified only as defective or non-defective, satisfactory or non-satisfactory, whereas in the variables sampling plans, the actual measurements are taken into account. As a result, the efficiency of the variables plans is increased considerably in the sense that, for the same degree of protection, it would require a smaller number of samples to be tested. In Indian Standards both attributes and variables plans have been used for ascertaining comformity. In those cases where the number of samples for test has to be decreased from practical considerations, interpretation on the basis of variables sampling plans is preferred. A further reduction in the amount of testing is also brought about, wherever appropriate, by testing a composite sample (obtained by mixing equal quantities of material from individual samples), for less important characteristics. However, for more important characteristics, the individual samples may be tested so as to obtain the requisite quality assurance. There are several Indian Standards where sampling plans have been given on such considerations such as the following :

IS : 560-1961	BHC, technical *(revised)*
IS : 670-1963	Specification for serge, worsted, dyed (superior) *(revised)*
IS : 779-1961	Specification for water meters (domestic type) *(revised)*
IS : 1606-1966	Specification for automobile lamps *(revised)*
IS : 2051-1962	Methods for sampling of leather footwear
IS : 2788-1964	Specification for gas mantles

Although the variables sampling plans and/or composite testing often reduce the sampling and testing cost to a considerable extent, yet these costs may still be uneconomical, particularly when the cost of material is high and the testing is expensive and time consuming. In such cases, an alternative procedure is to determine certain characteristics which are highly correlated with the destructive characteristics, but at the same time their testing should be simpler and less expensive. Efforts are being constantly made by ISI to examine such possibilities in the testing of various characteristics given in the material specifications.

Sampling of bulk materials like coal, gypsum, iron ores, manganese ores, etc., presents altogether different problems. These materials, unlike discrete items, consist of dissimilar particles of various sizes which differ appreciably in respect of quality characteristics, thereby making it difficult to collect representative samples for test. In such cases, it becomes necessary to conduct detailed investigations on a statistically designed basis with a view to collecting relevant information regarding the optimum weights of the gross sample, the optimum size of the increment, the method of reduction of the gross sample including that of coning and quartering, etc, for laying down rational sampling procedures. To give proper guidance in such cases, ISI has prepared a number of standards on the methods for sampling of these materials, and has more under preparation. An important feature of these sampling procedures is the subdivision of a lot into a number of sub-lots depending upon the weight of the lot and then drawing a representative sample from each sub-lot and testing them separately after reduction. This approach differs from the conventional one wherein one gross sample used to be tested and, therefore, no information as to the variation of the quality of the lot was available. According to the new procedure several test results will be obtained which would not only be self-checking but also provide an estimate of the variation in quality.

Certification of Quality :

To make quality goods available to consumers, ISI operates a Certification Marks Scheme under which licences are issued to manufacturers for affixing ISI mark on their products to guarantee that the quality is in accordance with the relevant Indian Standard. To enable that the marked products conform to the relevant Indian Standards, the manufacturer has to agree to follow a scheme of routine inspection and testing which is drawn up by ISI in consultation with the manufacturer. These inspection schemes are evolved on the basis of quality control principles and techniques. They include a sampling plan which takes into account the importance of the various characteristics, the number of tests that should be made at the various levels and strategic points of control and the criteria for determining the conformity of the control units to the relevant specification.

According to the inspection schemes, adequate records about the quality during production have also to be maintained by the licensees on suitable proformas prescribed for this purpose. From these records, the licensee can determine whether or not the production line is progressing satisfactorily ; if not, he can rectify the process so that the products conform to the relevant Indian Standard.

Indian Standards Institution also brings to the notice of the management of the licensees the benefits accruing from the use of quality control techniques and assists them in introducing these techniques into their plants. Recently ISI has also started its training programmes in statistical quality control for the benefit of its licensees in different sectors of industry. It is also interesting to mention here that the introduction of quality control techniques is making available systematic data on the capability of different manufacturing units for uniform production of quality goods conforming to Indian Standars. This information is most useful to ISI in revising Indian Standards from time to time to reflect the improved quality performance in the product.

RECENT DEVELOPMENTS ON THE CONSTRUCTION OF MUTUALLY ORTHOGONAL LATIN SQUARES

By

S. S. SHRIKHANDE

Summary. This paper deals with the recent developments on the construction of sets of mutually orthogonal latin squares (mols) with special reference to the falsity of the conjectures of Euler and MacNeish. Three methods have emerged in this direction and they depend upon the use of (*i*) incomplete block designs, (*ii*) difference sets and (*iii*) orthogonal mappings in groups [2, 3, 4, 5, 11, 17, 18, 19]. An attempt has been made in this paper which is as self-contained as possible to give all these methods and to indicate a large number of counter examples to both the above conjectures.

I. Introduction and Preliminary Results. A latin square of order u is a square array of order u in u distinct symbols such that each symbol occurs exactly once in every row and every column. Two latin squares of order u are said to be orthogonal, if on superposition, each symbol of the first square occurs exactly once with each symbol of the second square.

A set of latin squares all of the same order is said to be a set of mols if any two latin squares of the set are orthogonal. We exhibit a set of three mols of order 4 in four symbols 0, 1, 2, 3.

L_1	L_2	L_3
0 2 1 3	1 0 2 3	3 1 2 0
2 0 3 1	2 3 1 0	0 2 1 3
1 3 0 2	3 2 0 1	1 3 0 2
3 1 2 0	0 1 3 2	2 0 3 1

Consider a matrix $A=(a_{ij})$ of order $m \times u^2$ where each a_{ij} represents one of the symbols 1, 2,......, u. The matrix is called an orthogonal array $A(u^2, m, u\ 2)$ of size u^2, m constraints, u, levels strength 2 and index 1 if each 2-rowed sub-matrix of A contains each of possible u^2 ordered pairs exactly once[21].

Lemma I. The existence of m-2 mols of order u is equivalent to the existence of an array $A(u^2, m, u, 2)$.

Proof. Let $L_1, L_2, \ldots\ldots, L_{m-2}$ be set of mols of order u in symbols 1, 2,.. ..., u. Let $L_R\ (L_C)$ be a $u \times u$ array containing i in the ith row (column). Write each square as a row with u^2 entries such that the symbol in the cell (i, j) of the square occurs in position numbered $(i-1)u+j$. We then get a matrix with m rows and u^2 columns. We show that this forms an array $A(u^2, m, u, 2)$. It is obvious that in rows corresponding to L_R and L_C, every ordered pair occurs exactly once. From the properties of a latin square the same is true for rows corresponding to $L_R (L_C)$ and any latin square L_i. For rows corresponding to L_i and L_j, $i \neq j, = 1, 2, \ldots\ldots, m-2$, the result follows from the orthogonality of these latin squares. Thus the matrix obtained is an orthogonal array $A(u^2, m, u, 2)$.

Conversely given $A(u^2, m, u, 2)$, we can use any two rows to coordinatise the cells of a $u \times u$ square. Then corresponding to each one of the remaining m-2 rows form a square by putting in the cell (i, j) the symbol which occurs in that row in the corresponding position. We thus get a set of m-2 squares $L_1, L_2, \ldots\ldots, L_{m-2}$. From the definition of the orthogonal array it is easy to verify that each square is a latin square and further that any two of them are orthogonal.

Denote by $N(u)$ the maximum number of latin squares of order u such that any two are orthogonal. We then have the following result :

Lemma 2. $N(u) \leqslant u-1$

Proof. Suppose there exists a set of m-2 mols of order u. Consider the corresponding array $A(u^2, m, u, 2)$ given by Lemma 1. With respect to any given initial column of this array let n_i denote the number of columns having i coincidences with it, where two columns are said to have i coincidences if there are exactly i rows in which the two columns have the same symbol. Then from [1], we have

$$\sum_{0}^{m} n_i = u^2 - 1$$

$$\sum i n_i = m(u-1)$$

$$\sum i(i-1) n_i = 0$$

Then since the variance $V(i)$ of the set of real numbers i is non-negative, utilising the above relations we get

$$(u^2-1) \sum i^2 n_i \geqslant \sum (i n_i)^2$$

which gives

$$m \leqslant u+1$$

and hence

$$N(u) \leqslant u-1.$$

Remark. We note that if $m=u+1$, then $V(i)=0$ and hence $n_i=0$ for all i except $i=1$ and $n_i=u^2-1$ *i.e.*, any two columns of $A(u^2, u+1, u, 2)$ have exactly one coincidence.

We will call a set of $u-1$ mols of order u as a complete set of mols of order u.

Lemma 3. A complete set of mols of order u exists for every u which is a prime power.

Proof. Since u is a prime power there exists a finite field $GF(u)$ of order u. Let $\theta_0=0$, $\theta_1=1$, $\theta_2, \ldots\ldots, \theta_{u-1}$ be the elements of this field. Write these elements in a column and number the rows as $0, 1, \ldots\ldots \ldots u-1$ *i.e.*, θ_i occurs in row i. Consider the set of $u(u-1)$ transformations given by $y=\alpha x+\beta_j$ when α, $\beta \in (GF(u)$ and $\alpha \neq 0$. The initial column of elements of $GF(u)$ then gives rise to a matrix B of order $[u, u(u-1)]$. We note that the ordered pair $(\theta_i, \theta_j) i \neq j$ occurs exactly once in any two rows numbered l and k since the equations $\theta_i=\alpha\theta_l+\beta$ and $\theta_j=\alpha\theta_k+\beta$ have a unique solution. We further note that

$$B=(B_1 \;\; B_2 \ldots\ldots B_{u-1})$$

where B_i is a set of u columns generated from the initial column by taking $\alpha=\theta_i$ and all possible u values of β. It is then obvious that each row of B_i contains all the symbols of $GF(u)$ exactly once. Let B_0 be the square matrix of order u such that the ith column contains only the symbol θ_i ; $i=0, 1, \ldots$ $\ldots, u-1$. Then it is obvious that

$$(B_0 \quad B_1 \quad B_2 \ \ldots \ B_{u-1})$$

is an orthogonal array $A(u^2, u, u, 2)$, and that

$$\begin{pmatrix} B_0 & B_1 & B_2 & B_{u-1} \\ \theta_0 J' & \theta_1 J' & \theta_2 J' & \theta_{u-1} J' \end{pmatrix}$$

is an orthogonal array $A(u^2, u+1, u, 2)$ where J' is a row vector with u components each equal to 1. An appeal to Lemma 1 completes the proof.

Lemma 4. If Orthogonal arrays $A(u_1^2, m_1, u_1, 2)$ and $A(u_2^2, m_2, u_2, 2)$ exist and if $m=\min(m_1, m_2)$, then an orthogonal array $A(u_1^2 \, u_2^2, m, u_1u_2, 2)$ exists.

Proof. Let the symbols of the two arrays be $\alpha_1, \alpha_2, \ldots\ldots, \alpha_{u_1}$ and $\beta_1, \beta_2, \ldots\ldots, \beta_{u_2}$. Retain only the first m rows of both the arrays. Let each column of the first array containing the symbols $\alpha_{i_1}, \ldots\ldots \alpha_{i_m}$ in the first m positions be combined with each column of the second array containing $\beta_{j_1}, \ldots\ldots, \beta_{j_m}$ in these positions to form a new column containing the ordered pairs $(\alpha_{i_1}, \beta_{j_1}), \ldots\ldots, (\alpha_{i_m}, \beta_{j_m})$. We thus get a matrix of $u_1^2 \, u_2^2$ columns and m rows in u_1u_2 symbols (α_i, β_j), $i=1, 2, \ldots\ldots . u_1$; $j=1, 2, \ldots\ldots, u_2$. It is easy to verify that this matrix is the required array $A(u_1^2 \, u_2^2, m, u_1u_2, 2)$.

Corollary 1. For any two positive integers u_1 and u_2

$$N(u_1\, u_2) \geqslant \min\, (N(u_1).\ N(u_2)).$$

If $u = p_1^{n_1}\, p_2^{n_2}\, \ldots\ldots p_l^{n_l}$ is the prime power decomposition of u, then we define

$$n(u) = \min\, (p_1^{n_1},\, p_2^{n_2},\, \ldots\ldots,\, p_l^{n_l}) - 1.$$

Utilising Lemmas 3 and Corollary 1 above we have following result.

Corollary 2. $N(u) \geqslant n(u)$.

We note that

$N(u) = n(u) = u - 1$, if and only if u is a prime power,

and

$$n(u) = 1 \text{ if and only if } u = 4t + 2.$$

The existence of $n(u)$ mols of order u was proved by MacNeish[15], and Mann[16] by group-theoretic methods. MacNeish further conjectured that for all u, $N(u) = n(u)$. We will call this MacNeish's *conjecture* which implied that a complete set of mols of order u exists if and only if u is a prime power. On the other hand it also implied Euler's conjecture[7] that there does not exist a pair of mols of order $4t + 2$. It actually turns out that both these conjectures are invalid and that Euler's conjecture is true only for numbers 2 and 6. It may be mentioned that several attempts, necessarily incorrect, have been made in the past to prove Euler's conjecture *e.g.* Peterson[20], Wernicke[25], and MacNeish[15]. Levi[13] pointed out the inadmissibility of the arguments used by Peterson and MacNeish, Wernick's argument was shown to be fallacious by MacNeish himself[14]. The first proof of the validity of Euler's conjecture for the number 6 is due to Tarry[24].

2. Preliminary results on incomplete block designs. A balanced incomplete block design (bibd) is an arrangement of v symbols in b sub-sets of k distinct symbols $(k < v)$ satisfying the condition that any two distinct symbols occur together in exactly λ subsets. Using the usual terminology in statistics we call the symbols as treatments and subsets as blocks. It then follows that each treatment is replicated r times *i.e.* occurs in exactly r blocks and that the conditions

$$vr = bk$$

$$\lambda(v - 1) = r\,(k - 1) \tag{2.1}$$

must necessarily be satisfied. Fisher[8] showed that a further necessary condition for existence is $b \geqslant v$.

A bibd is called resolvable if the b blocks can be partitioned into r sets of n each such that each treatment occurs exactly once in each set. Necessarily then we have $v=nk$, $b=nr$. We shall be concerned only with bibd of index unity *i.e.* $\lambda=1$ and shall denote such a design by the notation (v, k) since the remaining parameters b and r can then be determined by (2.1).

Lemma 5. The existence of any one of the following three configurations (*i*) a complete set of mols of order u, (*ii*) a bibd with $v=u^2$, $b=u^2+u$, $r=u+1$, $k=u$, $\lambda=1$, and (*iii*) a bibd with $v=b=u^2+u+1$, $r=k=u+1$, $\lambda=1$ implies the existence of the remaining two.

Proof. We prove (*i*) → (*ii*) → (*iii*) → (*i*). Suppose (*i*) exists then there exists an array $A(u^2, u+1, u, 2)$ in symbols 1, 2, ..., u. Identify the symbol j in the i-th row with a treatment numbered $(i-1)$ $u+j$. Then we have u^2+u treatments. Take the columns of the array as the blocks of the design for these u^2+u treatments. Then each treatment occurs in u blocks and every block contains $u+1$ treatments. Further from the remark following lemma 2, it follows that any two blocks have exactly one treatment in common. Again it is obvious that the u^2+u treatments are partitioned into $u+1$ sets of u each such that any two treatments of the same set do not occur together, whereas two treatments coming from different sets occur together in exactly one block. It is now obvious that the dual[23] of this design is the resolvable bibd (*ii*).

Now suppose (*ii*) exists. Then it can be shown[23] that the design is resolvable. Let the treatments of (*ii*) be numbered 1, 2, ..., u^2 and let the $u+1$ sets of u blocks each forming a complete replication be R_1, R_2,......, R_{u+1}. Adjoin to each block of R_i a new treatment θ_i, $i=1, 2, \ldots\ldots, u+1$ and take an additional block containing $\theta_1, \theta_2, \ldots\ldots, \theta_{u+1}$. It is easy to see that this new design is exactly the design with parameters of (*iii*).

If (*iii*) exists, then it is easy to show that any two of its blocks have exactly one treatment in common. Take any block of the design and let this contain the treatments, say, $\theta_1, \theta_2, \ldots\ldots, \theta_{u+1}$. Consider the set of u remaining blocks which contain θ_i. This set contains u^2 treatments besides θ_i. From $\lambda=1$, it is obvious that this set of u^2 treatment contains none of the treatments $\theta_1, \theta_2, \ldots\ldots, \theta_{u+1}$. Since θ_i must occur together with every other treatment of (*iii*) exactly once, it is obvious that in this aggragate of u^2 treatments, all the treatments of (*iii*) except $\theta_1, \theta_2, \ldots\ldots, \theta_{u+1}$ must occur exactly once. It is now obvious that by omitting the block containing $\theta_1, \theta_2, \ldots\ldots, \theta_{u+1}$ and omitting these treatments we obtain a resolvable bibd with parameters of (*ii*) such that any two blocks of the same replication have no treatment in common and two blocks coming from different replications have exactly one treatment in common. Let the replications of the design (*ii*) thus obtained be $R_1, R_2, \ldots\ldots, R_{u+1}$. Number the blocks of each replication as 1, 2, ..., u in any arbitrary manner and let the treatments be numbered θ_1, θ_2, ..., θ_{u^2}. We form a $(u+1, u^2)$ martix in the following manner. Let the columns of the matrix by numbered θ_1, $\theta_2 \ldots \theta_u{}^2$. Corresponding to R_i we form the i-th row of the matrix in the following manner. If the treatment θ_m occurs in block numbered j of R_i, then we put the symbol j in the i-th row under the column numbered θ_m. Since the blocks of the same replication have no treatment in common, whereas the blocks of different replications have exactly one treatment in common, it is obvious that the matrix thus obtained is an orthogonal array $A(u^2, u+1, u, 2)$ which implies the existence of a complete set of mols of order u. The proof is thus complete.

Using Lemma 3 we have the following corollary.

Corollary. If u is a prime power, then each of the following configurations (*i*) a complete set of mols of order u, (*ii*) a resolvable bibd $(u^2 ; u)$, (*iii*) a bibd $(u^2+u+1 ; u+1)$ exists.

We now consider a generalization of a bibd with $\lambda=1$. An arrangement of v treatments in b blocks will be called a pairwise balanced design of index unity and type $(v: k_1, k_2, .., k_m)$ if each block contains either $k_1, k_2, \ldots k_m$ distinct treatments $(k_i < v, k_i \neq k_j)$ and every pair of distinct treatments occurs in exactly one block of the design. If the number of treatments containing k_i treatments is b_i, then obviously

$$b = \sum_i b_i, \quad v(v-1) = \sum_i b_i k_i (k_{i-1})$$

Consider a pairwise balanced design (D) of index unity and type

$$(v: k_1, k_2, \ldots, k_m).$$

The subdesign (D_i) formed by blocks of size k_i will be called the i-th equiblock component of (D), $i = 1, 2, \ldots, m$.

A subset of v blocks of (D_i) will be said to be of type I if every treatment occurs in the subset exactly k_i times. As pointed out by Konig [12] we can arrange the treatments within the blocks of the subset in such a way that every treatment occurs exactly once in every position. If the v blocks are written as columns, each treatment occurs exactly once in every row. When so written the blocks will be said to be in the standard form.

A subset of v/k_i blocks of (D_i) will be said to be of type II if every treatment occurs exactly once in the subset *i.e.* the subset is a complete replication of all the treatments. The component (D_i) will be said to be separable if its blocks can be divided into subsets of type I or type II. Both types may occur in (D_i) at the same time. The design (D) is said to be separable if each (D_i) is separable.

The set of equiblock components (D_1), (D_2),..., (D_e) is said to form a clear set, if any two blocks coming from D_i and D_j are disjoint whether $i=j$ or $i \neq j$. Obviously a necessary condition for this is

$$\sum_1^e b_i k_i \leqslant v.$$

We now prove a crucial lemma which is useful in disproving the conjectures of MacNeish and Euler.

Lemma 6. Suppose there exists a set Σ of $m-1$ mols of order k. Then we can construct a $[m, k(k-1)]$ matrix P, whose elements are the integers 1, 2,..., k and such that (*i*) any ordered pair (i, j), $i \neq j$ occurs as a column exactly once in any two rowed submatrix of P, (*ii*) P can be subdivided into $k-1$ submatrices P_1, P_2,..., P_{k-1} or order (m, k) such that in each row of P_i each of the symbols 1, 2,..., k occurs exactly once.

Proof. Renaming the symbols if necessary, it is obvious that the set Σ can be taken in the standard form in which the first row of each latin square contains the symbols 1, 2,..., k in that order. We now adjoin to the set Σ a (k, k) square containing the symbol i in each position in the ith column. We now write the elements of each square in a row such that the symbol in the ith row and jth column occupies the nth position where $n=(i-1)k+j$. It is then obvious that we get an orthogonal array $A(k^2, m, k, 2)$ in which the first m columns account for the pair (i, i) in any rows. By ommitting these columns, we obviously get the matrix P having the property (*i*). If P_i is the submatrix of P formed by columns $(i-1)k+1, j=1$, 2,..., k, then the property (*ii*) follows from the fact that each of the symbols 1, 2,...,k appears exactly once in any given row of latin squares belonging to Σ.

Let Y be a column of k distinct treatments chosen from the set t_1, t ,..., t_v,then we denote by $P(Y)$ the $[m, k(k-1)]$ matrix obtained from P by replacing the symbol i by the treatment occurring in the ith position in Y. A similar meaning will be assigned to $P_i(Y)$ and $\pi_{ij}(Y)$ where π_{ij} denotes the jth column of P_i. Clearly every treatment of Y occurs once in every row of $P_i(Y)$ and if t_a and t_b are any two elements of Y then the ordered pair (t_a, t_b) occurs as a column exactly once in any two rowed submatrix of $P(Y)$.

3. Use of incomplete block designs in constructing mols. We prove two theorems.

Theorem 1. [4] Let there exist a pairwise balanced design (D) of index unity of type $(v: k_1, k_2, \ldots k_l)$ and suppose there exist m_i-1 mols of order k_i. If

$$m=min(m_1, m_2, \ldots m_l)$$

then $N(v) \geqslant (m-2)$. If the design (D) is separable then $N(v) \geqslant m-1$.

Proof. Let the treatments of the design be $t_1, t_2,\ldots,t_v$ and let the blocks of the design (written as columns) belonging to the equiblock component (D_i) be $Y_{i_1},\ldots,Y_{i_{b_i}}$ $(i=1, 2,\ldots l)$. Define the $k_i \times b_i$ matrix D_i by

$$D_i=(Y_{i_1}, Y_{i_1},\ldots,Y_{i_{b_i}})$$

Let P_i be the matrix of order $[m_i, k_i(k_i-1)]$ defined in the previous lemma, the elements of P_i being the symbols 1, 2,..., k_i.

Let P_{i_c}, $c=1, 2,\ldots k_i-1$ be the submatrices of P_i such that each row of P_{i_c} contains the symbols $1, 2,\ldots k_i$ exactly once.

Put $$P_i(D_i)=P_i(Y_{i_1})\ldots P_i(Y_{i_{b_i}})$$

Then $P_i(D_i)$ is of order $[m_i, b_i\, k_i(k_i-1)]$ and if t_a and t_b are any two treatments occurring in the same block of (D_i) then the ordered pair (t_a, t_b) occurs exactly once in any two rowed submatrices of $P_i(D_i)$. Let $\triangle_i$ be the matrix obtained from $P_i(D_i)$ by retaining only the first m rows and let

$$\triangle=(\triangle_1\ \ \triangle_2\cdots\triangle_l)$$

Then $\triangle$ has $\Sigma b_i\, k_i(k_i-1)=v(v-1)$ columns and it follows that any ordered pair of distinct elements from the set $t_1, t_2,\ldots t_v$ occurs as a column exactly once in any two rowed submatrix of $\triangle$. Let $\triangle_0$ be a (m, v) matrix whose i-th column contains the symbol t_i in every position, $i=1, 2,\ldots v$. Then the matrix $(\triangle_0\ \ \triangle)$ is obviously an array $A(v^2, m, v, 2)$ and hence $N(v)\geqslant m-2$.

To prove the second part consider each component (D_i) of the separable design (D). If (D_{i_j}) is any set of v blocks of (D_i) of type I, then as mentioned before (D_{i_j}) can be taken in the standard form so that each row of (D_{i_j}) contains all the v treatments exactly once. If π_{i_c} is any column of P_i, then defining $\pi_{i_c}(D_{i_j})$ in the obvious manner, it is obvious that all the v treatments occur exactly once in any row of $\pi_{i_c}(D_{i_j})$. Similarly if we have a set of $\frac{v}{k_i}=n$ blocks (D^*_{ij}) of type II and $P_i=(\ldots\ldots, P_{i_c}\ldots\ldots)$, $c=1, 2\ldots$, $k-1$ where each P_{i_c} contains all the k_i symbols exactly once then it is obvious that $P_{i_c}(D^*_{i_j})$ contains v columns with the same property, *i.e.*, each row of $P_{i_c}(D^*_{i_j})$contains all the v treatments exactly once. It is now obvious that the matrix $(\triangle_0\ \ \triangle)$ defined as before is an array $A(v^2, m, v, 2)$ which is resolvable, *i.e.*,

$$\triangle=(\triangle_0\ \ \triangle^{(1)}\cdots\ \triangle^{(v-1)})$$

where each $\triangle_0$ or $\triangle^{(i)}$ represents a set of v columns such that in each set every treatment occurs exactly once in each row. If α_i represents a row vector with v components each of which is t_i then obviously

$$\begin{pmatrix}\triangle_0 & \triangle^{(1)}\ldots\ldots\triangle^{(v-1)}\\ \alpha_0 & \alpha_1\ \ \ldots\ldots\alpha_{(v-1)}\end{pmatrix}$$

is an array $(v^2, m+1, v, 2)$ and hence $N(v)\geqslant m-1$.

Theorem 2. [5]. Let there exist a design (D) of index unity and type $(v\,; k_1. k_2 \ldots, k_l)$ such that the set of equiblock components (D_1). $(D_2) \ldots$, (D_e), $e < l$ is a clear set. If there exist $m_i - 1$ moles of order k_i, and if

$$m^* = min(m_1 + 1, \ldots ,m_e + 1\ m_{e+1}, \ldots, m_l)$$

then
$$N(v) \geqslant m^* - 2.$$

Proof. Let A_{i_j} be an orthogonal array with $m_i + 1$ rows in symbols of Y_{i_j} where Y_{i_j} is the j-th block of (D_i), $i = 1, 2 \ldots, e$. Put

$$A_i = (A_{i_1}, \ldots, A_{ib_i})$$

Let $\triangle_i$ be the $(m^*, b_i k_i^2)$ matrix obtained from A_i by retaining only the first m^* rows, and let

$$\triangle^{(i)} = (\triangle_1\ \triangle_2 \ldots \ldots \triangle_e)$$

Then $\triangle^{(1)}$ is of order $(m^*, \sum_1^e b_i k_i^2)$ and has the property that if t_b and t_d are any two treatments occurring in a block of (D_i), $i = 1, 2, \ldots, e$, then the ordered pair (t_c, t_d) occurs as a column in any two rowed submatrix of $\triangle^{(1)}$ whether $c = d$, $c \neq d$.

Consider the matrix P_i of order $(m_i, k_i (k_i - 1))$ defined in Lemma 6 for $i = e + 1, \ldots, l$ and let $\triangle_i$ be the matrix obtained from $P_i(D_i)$ by retaining only the first m^* rows. Then

$$\triangle^{(2)} = (\triangle_{e+1}\ \triangle_{e+2} \cdots \triangle_l)$$

has the property that if t_a and t_b are any two distinct treatments contained in any block of $(D_{e+1}, \ldots, D_l)$, then the ordered pair (t_a, t_b) occurs exactly once in any two rowed submatrix of $\triangle^{(2)}$. Let $\triangle^{(3)}$ be the (m^*, v_2) matrix whose n-th column contains t_n in every position, where t_n is any one of the $v_2 = v - \Sigma b_i k_i$ treatments not occuring in $(D_1), \ldots, (D_e)$. Then $(\triangle^{(1)}\ \triangle^{(2)}\ \triangle^{(3)})$ is obviously an array $A(v^2, m^*, v, 2)$ and hence $N(v) \geqslant m^* - 2$.

We now illustrate the above theorem by giving some examples.

Since 7 is a prime there exists a resolvable bibd (49; 7). Adjoin a new treatment θ to each block of a replication of this design. We obviously get a pairwise balanced design of type (50 ; 7, 8). Application of Theorem 1 then shows that $N(50) \geqslant 5$ whereas $n(50) = 1$. This is a counter example to both Euler's and MacNeish's conjecture.

For a bibd with $k = 5, \lambda = 1$, v is of the form $20t + 1$ or $20 + 5$. It is known [9, 10] that these designs exists for all value of t. Consider treatments θ_1, θ_2, θ_3 from a design $(v\,; 5)$ such that all the three treatments do not occur in the same block of the design. Then by omitting these treatments from the design we get a design of the type $(v-3;3,4,5)$ where the 3 blocks of size 3 form a clear set. Theorem 2 now gives,

$$N(v) \geqslant 2 \text{ if } v=20t+2.$$

which provides an infinity of counter examples to Euler's conjecture.

As an immediate consequence of Corollary 1 of Lemma 4 we have the following theorem.

Theorem 3. If there exist 2 mols of any order $4t+2$, then there exist at least 2 mols of any order which is an odd multiple of $4t+2$.

We now state without proof the following theorem [5] which plays a crucial role in the complete disproof of Euler's conjecture for all $v > 6$.

Theorem 4. If $k \leqslant N(m)+1$, then

(1) $N(km+1) \geqslant \min[\, N(k), N(k+1), 1+N(m)] - 1$

(2) $N(km+x) \geqslant \min(N(k), N(k+1), 1+N(m), 1+N(x)) - 1$ if $1 < x < m$.

We conclude this section by stating two asymptotic results on the behaviour of $N(v)$.

Theorem 5 [6]. There exists a number v_0 such that for all $v > v_0$ we have

$$N(v) > \tfrac{1}{3}v^{1/91}$$

Theorem 6 [22]. To each number $c < \frac{1}{42}$, there exists an integer $r_0 = v_0(c)$ such that for all $v > v_0$

$$N(v) > v^c$$

The proofs of the above two theorems are purely number theoretic and they do not involve any combinatorial insight into the problem and are based on the inequalities in Theorem 4.

4. Use of the method of differences. Let $0, 1, 2, \ldots, n-1$ be the elements of the ring R of residue classes (mod. n). We consider matrices whose elements belong either to R or to the set X of m indefinites $x_1, x_2, \ldots, x_m$. We say that the difference associated with the ordered pair (i, j) where i and j belong to R is c where $c \equiv i-j$ (mod, n), $0 \leqslant c < n$. Conversely to each element c of R there corresponds n ordered pairs (i, j) which have c as their associated difference. If (i, j) is one such pair then the other pairs are

$$(i+\theta, j+\theta) \text{ where } \theta = 0, 1, \ldots n-1 \text{ where } i+\theta \text{ and } j+\theta$$

are reduced (mod n). The ordered pair (i, j) both members of which belong to R will be called an R-pair. For i in R and x_j in X the ordered pair (i, j)

is called an $R-X$ pair and the difference associated with it is defined to be x_j. If θ is any element of R we formally define $0+x_j=x_j$. With this definition corresponding to any x_j, there are n $R-X$ pairs the difference associated with which is x_j. These pairs are of course the pairs

$$(i, x_j), \ i=0, 1, \ldots, n-1.$$

We similarly define $X-R$ pairs. The difference associated with $X-R$ pair (x_j, i) is x_j.

Theorem 7 [5]. If m is odd there exist at least 2 mols of order $3m+1$. Taking $m=4t+3$ this implies the existence of 2 mols of all ordered $12t+10$.

Proof. Consider the (4, 4m) matrix A_0 given below whose elements belong to R the residue classes mod $(2m+1)$ or X the set of indefinites $x_1, x_2, \ldots, x_m$.

$$A_0=\begin{bmatrix} 0 & 0\ldots0 & 1 & 2\ldots m & 2m & 2m-1\ldots m+1 & x_1 & x_2\ldots x_m \\ 1 & 2\ldots m & 0 & 0\ldots0 & x_1 & x_2\ldots x_m & 2m & 2m+1\ldots m+1 \\ 2m & 2m-1\ldots m+1 & x_1 & x_1\ldots x_m & 0 & 0\ldots0 & 1 & 2\ldots m \\ x_1 & x_2\ldots x_m & 2m & 2m-1\ldots m+1 & 1 & 2\ldots m & 0 & 0\ldots0 \end{bmatrix}$$

We note that the $4m$ pairs occurring as columns in any two rowed submatrix of A_0 are $2m$, R—pairs the differences associated with which are all the nonnull elements of R, m, $R-X$ pairs the differences associated with which are all the elements of X and m $X-R$ pairs the differences associated with which are all the elements of X. Let A_θ be the matrix obtained from A_0 by adding θ, $0\leqslant\theta\leqslant 2m$, to every element of A_0 and reducing (mod $2m+1$); $x_j+\theta$ being considered as x_j.

Put

$$A=(A_0 \ A_1 \ \ldots A_{2m})$$

Then it is evident that in any two rowed submatrix of A, any R pair formed by two distinct elements of R, or any $R-X$ or $X-R$ pair occurs exactly once. Let A^* be an orthogonal array $(m^2, 4, m, 2)$ corresponding to 2 mols of order m in symbols X. Such an array always existss ince $N(m)\geqslant n(m)\geqslant 2$ as n is odd. Further let E be a $(4, 2m+1)$ matrix whose ith column contains i in each position, $0\leqslant i\leqslant 2m$.

Then

$$\triangle=(E \ A \ A^*)$$

is an orthogonal array $[(3m+1)^2, 4, 3m+1, 2]$, which proves the result.

As a corollary by taking $t=0$, 1, 2, 3, 4, 8 repectively we obtain

$$N(v)\geqslant 2 \ \text{ for } v=10, 22, 34, 46, 58, 106.$$

5. Use of orthogonal mappings in a group. In this section we describe briefly the technique developed by Bose, Chakravarti and Knuth [2] for constructing sets of mols.

Let G be a finite group of order n and α be a (1, 1) mappings of G onto itself, the image of x in G under α being denoted by $\alpha(x)$.

Let M denote the set of all (1, 1) maps of G onto itself. Then the order of M is $n!$ We shall use Greek letters for elements of M except that I shall stand for the identity map. To each α there corresponds a unique inverse map α^{-1} in M such that if $\alpha(x)=c$ then $\alpha^{-1}(c)=x$. We denote by $\beta\alpha$ the map for which $\beta\alpha(x)=\beta(\alpha(x))$. It is clear that $\beta\alpha$ is in M and the associative law $\alpha\beta(\gamma)=\alpha(\beta\gamma)$ holds.

Two maps α and β in M are said to be orthogonal if the equation

$$(\alpha x)(\beta x)^{-1}=c$$

has a unique solution x in G for any c in G.

Consider a (n, n) square. We can make a (1, 1) correspondence between the rows of the square and the elements of G. Similarly we can make a (1, 1) correspondence between the columns of the square and the elements of G. The cell of the square corresponding to row x and column y is denoted by the cell (x, y). We then have the following results.

Theorem 8. If in the cell (x, y) of an (n, n) square we put the element $(\alpha x)y$ of G we get a latin square $L(\alpha)$ where α is in M. Proof is obvious.

Theorem 9. The necessary and sufficient condition for two Latin squares $L(\alpha)$ and $L(\beta)$ to be orthogonal is that the maps α and β in M are orthogonal.

Proof. If α and β are orthogonal maps in M to prove that $L(\alpha)$ and $L(\beta)$ are orthogonal it is enough to show that for any pair of elements u, v, in G the simultaneous equations

$$(\alpha x)y=u,\ (\beta x)y=v \tag{5.1}$$

have a unique solution (x, y) in G. The above equation is simply

$$(\alpha x)(\beta x)^{-1}=uv^{-1} \tag{5.2}$$

and since α and β are orthogonal map equation (5.2) has a unique solution for x and then (5.1) gives

$$y=(\alpha x)^{-1}\ u=(\beta x)^{-1}v.$$

Conversely if L (α) and L (β) are orthogonal, then equations (1) have have a unique solution and hence (2) determines a unique value of x for any pair u, v in G. This implies that the maps α and β are orthogonal.

In view of the above two theorems the search for a maximal set of mols. with elements in G reduces of finding a maximal set of mutually orthogonal maps in G.

Let G be an abelian group of order 12, whose elements are vectors (a,b), where a is a residue class (mod 2) and b is a residue class (mod 6) and the law of composition (addition) is

$$(a_1, b_1)+(a_2, b_2)=(a, b)$$

$$\text{where } a=a_1+a_2 \pmod 2,\ b=b_1+b_2 \pmod 6.$$

Utilising the existence of a symmetric bibd with $v=b=11$, $r=k=5$, $\lambda=2$ and making use of a high speed computor Bose Chakravarti and Knuth[2] were able to discover two maximal sets of 5 mutually orthogonal maps I, α_1, α_2, α_3, α_4 and hence a set of 5 mols of order 12. One such set is reproduced below where the vector (a_i, b_i) is written a_ib_i for short, *i. e.*, 14 instead of (1,4)

Ix_j	00	01	02	03	04	05	10	11	12	13	14	15
α_1x_j	00	02	01	14	13	15	03	05	04	11	10	12
α_2x_j	00	14	04	12	02	10	13	01	11	05	15	03
α_3x_j	00	05	13	02	14	11	12	04	15	01	03	10
a_4x_j	00	10	05	11	15	04	02	12	14	03	01	13

Using somewhat similar methods and also utilising a high speed computer, Mendelson, Dulmage, and Johnson[11] have also obtained a set of 5 mols of order 12 and they conjecture that a set of $2p$—1 mols of order $4p$ exists for any prime p.

REFERENCES

1. R.C. Bose and K.A. Bush (1952) : Orthogonal Arrays of strength two and three. Ann. Math. Statist. 23, 508-524.

2. R.C. Bose, I.M. Chakravarty and E. D. Knuth (1960) : On methods of constructing sets of mutually orthogonal Latin squares using a computer. 1. Technometrics 2, 507-516.

3. R.C. Bose and S.S. Shrikhande (1959) : On the falsity of Euler's conjecture about the non-existence of two mutually orthogonal Latin squares of order 4t+2. Proc. Nat. Acad. Sci. U S.A., 45, 734-739.

4. R.C. Bose and S S. Shrikhande (1960) : On the construction of sets of mutually orthogonal Latin squares and the falsity of a conjecture of Euler. Trans. Amer. Math. Soc. 95, 191-209.

5. R.C. Bose, S S. Shrikhande and E.T. Parker (1960) : Further results on the construction of mutually orthogonal Latin squares and the falsity of Euler's conjecture Canad. J. Math. 12, 189-203.

6. S. Chowla, P. Erdos and E.G. Straus (1960) : On the maximal number of pairwise orthogonal Latin squares of a given order. Canad, J. Math. 12, 204-208.

7. L. Euler (1782) : Recherches sur une nouvelle especes des quarres magique. Verh, Zeeuwsch Genoot, Wetenschappen. 9, 85-239.

8. R.A. Fisher (1940) : An examination of the different possible solutions of a problem in incomplete block configurations. Ann. Eug. 10, 52-75.

9. H. Hanani (1961) : The existence and construction of balanced incomplete block designs. Ann. Math. Statist. 32, 361-386.

10. H. Hanani (1965) : A balanced incomplete block design, Ann. Math. Statist. 36, 711.

11. Diane M. Johnson, A.L. Dulmage and N.S. Mendelson (1961) : Orthogomorphisms of groups and orthogonal Latin Squares, Canad. J. Math. 13, 356-372.

12. D. Konig (1950) : Theorie der Endlichen und unendlichen Graphen. Chelsea, New York.

13. F.W. Levi (1942) : *Finite Geometrical Systems*. University of Calcutta.

14. H.F. MacNeish (1921) : Das problem der 36 offiziere. Jber. Deutsch. Math. Verein 30, 151-153.

15. H.F. MacNeish (1922) : Euler's squares. Anu. Math. Statist. 23, 221-227.

16. H.B. Mann (1942) : The construction of orthogonal Latin squares, Ann. Math. Statist. 13, 418-423.

17. P.K. Menon (1961) : Methods of constructing two mutually orthogonal Latin Squares of order $3n+1$. Sankhya series A. 23, 281-282.

18. E.T. Parker (1959) : Construction of some sets of mutually orthogonal Latin squares. Proc. Amer. Math. Soc. 10, 946-949.

19. E.T. Parker (1959) : Orthogonal Latin Squares. Proc. Nat. Acad. Sc. U.S.A. 45, 859-862.

20. J. Peterson. (1901-1902) : Les 36 officiers. Ann. of Math. 413-417.

21. C.R. Rao (1947) : Factorial experiments derivable from combinatorial arrangement of arrays. J. Roy. Stat. Soc. Suppl. 9, 128-139.

22. K. Rogers (1964) : A note on orthogonal *Latin Squares, Pacific J. Math. 14*, 1395-1397.

23. S.S. Shrikhande (1952) : On the dual of some balanced incomplete block designs Biometrics, 8, 66-72.

24. G. Tarry (1960) : Le probleme de 36 officiers. Ass. France Av. Sci. 170-203.

25. P. Wernicke (1910) : Das problem der 36 offiziere Jber. Deutsch Math verein, 19, 264-267.

DOUBLE SAMPLING AND ITS APPLICATION IN AGRICULTURE

By

D. SINGH*

I. Introduction.

Demand of statistics on various facets of economy in all the countries is increasing at a fast rate. In developing countries basic data are needed for formulating developmental plans and for subsequent assessment of their progress, while in the developed countries such data are being accumulated for devising new techniques of accelerating the production. The process of planning for the economic and social advancement consists in rationally allocating the resources of the country first to the different sectors of the economy such as agriculture, industry, education, social services, etc. and then within each of these sectors to the administrative divisions such as states, district, community development block, village, etc. For this purpose it becomes necessary to collect relevant information regarding the different sectors of economy for each of the administrative divisions. For example, data on various aspects of agriculture are at present needed for each of the community development blocks in India for fixing the target of agricultural production and for assessing the progress of various developmental programmes being conducted therein.

Of all the sectors of economy, it is most difficult to obtain reliable data on agriculture. It is more so in the developing countries where agriculture is the mainstay of the entire economy. Most of these developing countries lack necessary resources and personnel to collect reliable data on agriculture on frequent intervals, which are needed for planning purposes. The Food and Agriculture Organisation of the United Nations, to fill up this serious gap, has recently provided different types of aids to organize agricultural census in these countries. A few countries, however, have taken advantage of such aids provided by the F.A.O. and other International Agencies. The countries which have conducted such agricultural census particularly on complete enumeration basis are now in a better position to utilise reliable data for planning agricultural development on scientific lines.

The census is a costly and complicated operation and it cannot, therefore, be operated on a very frequent interval. Even the developed countries, which possess plently of resources, plan agricultural census at an interval of more than five years. In the changing economy, information is to be obtained for the intercensal period by planning scientific surveys which are economical and operationally feasible. This necessity has led to the considerable development in the application of modern sampling technique based on probability theory.

---

*Institute of Agricultural Research Statistics, New Delhi.

A most significant development in theory of sampling has recently taken place in marshalling the entire information available on the character under study by using it either at the stage of designing or at the stage of estimation depending upon in what form such information is available. Basic idea behind this development is that the available resources at the disposal of the experimenter should be used with utmost care and maximum information should be obtained for a set total expenditure. Use of stratification, selection of sampling units with varying probabilities, method of control selection, rotation sampling, etc. are some of the examples of using the relevant information on the units of population at the stage of designing the survey, while ratio and regression methods of estimation are the examples of using the information at the stage of estimation.

Techniques mentioned above depend upon the possession of the advance information, about auxiliary variates x_i's closely related to the character (say, y_i) under study. When such information is lacking it is sometimes relatively cheap to take a large preliminary sample in which x_i's alone are measured. The objective of this sample is to obtain good estimate of X or of the frequency distribution of x_i's.

This procedure implies that in a survey where main function is to make estimate for variate y_i, it may pay to devote part of their resources to the primary sample, although this means that the size of the sample in the main survey on y_i must be reduced. The technique is known as double sampling or two-phased sampling. It is obvious that the technique will pay only if the gain in precision more than offsets the loss in precision due to the reduction in the size of the main sample.

2. Review.

Watson (1937) was the first experimenter who took advantage of the technique of using information on auxiliary variable to improve efficiency of the estimate of the variable in which he was interested. His problem was to estimate the area of leaves of field crops but he found it tedious and time consuming work to measure area of large number of leaves. He, therefore, took the weight of all leaves and then measured area as well as weight for a sub-sample of leaves and using the regression technique he estimated the area of leaves of field crops.

The concept of double sampling was first introduced by Neyman (1938). He felt the need of double sampling technique while examining the problem of stratification. The technique of stratification no doubt improves the estimate of the character under study provided it is possible to choose an appropriate character highly correlated with the character under study for the purpose of stratification. If information on such character is not readily available but it could be easily collected cheaply it is worthwhile to spend a part of the resources for collection of such information. If a character under study (y_i, say) is costly to enquire about but a highly correlated and cheaper character (x) is available, it may be beneficial to draw a bigger random sample to collect information on x and to stratify the population on that basis.

Snedecor and King (1942) have mentioned the application of double sampling method, by Goodman for determination of corn yield. Goodman observed that it was easier and much cheaper to count the number of ears of corn in a given unit area than to harvest the yield and obtain the dry weight of kernels. The high cost of making dry weight determination suggested the possibility of double sampling in which ears would be counted and measured in many fields but harvested in only a portion of these. From some earlier experiments it was known that the product of length times diameter of ear is highly correlated with the dry weight of kernels ; hence the dry weight of kernels per ear could be estimated by the formula

$$E_y = \overline{y}_s + b(\overline{x}_L - \overline{x}_s)$$

where $\overline{y}_s$ = average dry weight of kernels for the small harvested sample Y_s,

$\overline{x}_L$ = average size (length × diameter) of ear for the larger sample in which measurements only are taken,

$\overline{x}_s$ = average size of ear for the small sample,

and

b = regression coefficient for the regression of y on x.

The variance of the estimate is given by the formula

$$V(E_y) = \sigma_y^2 \left[(1 - \rho^2) \frac{1}{n} + \frac{\rho^2}{N} \right]$$

where ρ = correlation between the size of the ear and dry weight of kernels in the population,

n = size of the small sample,

and

N = size of the large sample.

An interesting application of technique of eye appraisal (as auxiliary variable) has been mentioned by Yates (1960) for estimation of the total volume of certain species of timber in Great Britain. In this study, a complete survey of the area was made. Each block of wood was visited and divided by eye inspection into areas uniform for descriptive purpose. These areas 'stands' were numbered and demarcated on the map and eye estimates were made of each species of timber in each 'stand'. Independent estimates of volumes were obtained by actual measurement of the trees in a small number of sampled plots located on a grid pattern across the area. On comparing the estimates of volume by measurement of the sampled plot with the eye estimates for stands in which sampled plots lay, the eye estimates were found to have high correlation with those estimated on the basis of actual measurement although there was an indication of negative bias in the eye estimates. The problem was whether subjective method of estimate of volume of timber by eye appraisal which gave negative bias, could be adopted as a procedure for estimation of volume of timbers. Yates observed that as long as there was real correlation between the eye estimates and those obtained on the basis of actual measurement, data based on eye estimation method could be used for improving the efficiencies of the estimate. He, therefore, recommended the use of eye estimate for the whole area and actual physical measurement for a small sample and then calculating the regression of the actual measured value on the corresponding eye estimate values. The regression estimate recommended by him takes the following form :—

$$Y_E = \bar{y}_s + b(\bar{x}_a - \bar{x}_s)$$

where $\bar{y}_s$ = the estimate of volume per unit of area based on actual measurement,

$\bar{x}_s$ = the corresponding eye estimate,

$\bar{x}_a$ = the eye estimate for the whole area,

and b is the regression coefficient of actual measurement on eye estimate based on small sample. The method where eye estimate is not used can be simply obtained by putting $b = 0$. This will also be so if eye estimate provides no information about the volumes. The robustness of the method lies in that it uses all possible information contained in the eye estimates in addition to that already provided by the sampled plot volumes obtained on actual measurement. He found on actual calculation that this procedure of using data based on eye appraisal for estimation of volume of timber of a particular species increased the efficiency by about 33 per cent over that where eye estimate was not used.

Cochran (1963) has suggested an application of eye estimate in conjunction with actual counting for estimating the rat population in an area. A rat expert may make a quick eye estimate of the number of rats in each block of a city area and then determine by trapping the actual number of rats in each of a random sample of blocks.

Sukhatme and Koshal (1959) worked out the ratio estimate by using double sampling technique when the design was multi-stage. They illustrated the efficiency of the double sampling procedure with the help of crop-cutting data on paddy conducted in Mansura district of Egypt. They used the data relating to weight of grains and straw together as auxiliary variable, x and dry grain alone as variable, y. Problem was to estimate the yield of dry grain. But the procedure of harvesting the produce, threshing it and obtaining the dry grain was much more costly than taking the record of grain and straw together on the date of harvesting. It was, therefore, suggested that in a large sample only records of grain and straw together might be taken while in a sub sample along with grain and straw, data on dry grain should also be obtained. They observed that under certain situation the gain in efficiency may be more than hundred per cent.

Singh & Singh (1965) considered a sampling procedure involving application of double sampling for stratification on several successive occasions. The application of the procedure was demonstrated in a survey conducted for estimating the total number of palms, area under coconut crop and yield per acre in the State of Assam. The initial sampling design adopted for the survey was one of stratified multi-stage random sampling, with sub-divisions of districts as strata and villages as first stage units. After two rounds of this survey were over it was observed that in more than one-third of the villages in Assam coconut crop was not grown and that inclusion of such villages in the sample increased the standard error of the sample estimate considerably. Consequently, it was decided to adopt two phase multi-stage sampling design from the third round onwards. According to the procedure a large sample was taken for stratifying the villages into two groups, those having coconut cultivation and those not having coconut cultivation. Then a sub-sample was taken from the group of villages reported to have coconut cultivation for collection of other data like number of palms, area, yield, etc.

The procedure of double sampling has been widely used in crop estimation surveys in India (1951). For estimation of yield of all the principal crops random crop-cutting experiments are annually planned in India. According to the procedure, a small random plot is located in the sampled field growing a crop and the crop of the plot is harvested on a fixed day. Since, on the day of harvest, the grains contain moisture, it becomes necessary to plan an elaborate procedure for obtaining driage ratio. To reduce the expenditure and labour, driage experiments are conducted in only a sub-sample of plots selected for crop-cutting surveys. The yield on the harvest day is multiplied by the driage ratio to obtain the dry yield.

3. An Application

The technique of the random sampling crop-cutting survey has now been recognised as the most reliable method for estimating the yield of crops not only in India but also in most parts of the world. The method possesses objectivity and simplicity in procedure. The major difficulty in strictly following the procedure has, however, been in finding suitable field organisation to complete the crop-cutting work in time. Since a particular crop gets harvested in a short time, the entire crop-cutting work has to be completed within that period. This has caused considerable difficulty in finding adequately trained field staff to complete the work in time. Consequently, the number of crop-cutting experiments planned on a crop has been limited in each area in proportion to the number of staff available in that area and the estimate of crop yields with reasonable degree of precision is at present at the most available at the regional level. For example in India, such estimates based on random crop-cutting experiments are available only at the district level. The Statisticians in India, have, therefore, so far expressed their inability to provide estimates of crop yields for smaller area than the district and the food administrators are content with this type of estimate. Now, there is a demand for reliable estimates for smaller areas like community development blocks or even for the individual villages. This need has arisen due to the planning of agricultural development for increasing production rapidly and for evolving a proper food distribution policy. Since community development blocks have become the unit of planning it is necessary that reliable data needed for formulating scientific plan and for assessing their progress should be available. The planners should also have an opportunity of examining the achievements as a result of various plan efforts made in the area. The key indicator of the achievements of such plan efforts is the agricultural production. It has, therefore, become obligatory on the part of the planners not only to concentrate on plan efforts, but also to know the results so that if it is considered necessary, modification in the plan activities can be made in the area. Target of procurement of agricultural produce in an area will also depend upon the per capita production. For rational procurement policy it is desirable to fix up the target of procurement according to the capacity of the blocks or villages. This will need knowledge on the production level for such areas. The new strategy of agricultural development programmes concerns at present with the spread of high yielding varieties of crops. Since high yielding varieties of crops require special type of facilities, planning of such programme will have to be based on principles of planning for small areas like a block or a group of blocks. Assessment of progress of such programmes can be made only if reliable information on yield is made available for smaller areas than the district.

Once it is admitted that reliable estimates for small areas like blocks will be useful for planning agricultural development and for framing food administrative policy, it becomes essential to evolve a method which can provide reliable estimate of crop production for such small areas. If the simple crop-cutting approach is to be adopted directly for this purpose the present number of crop-cutting experiments will have to be increased manifold. As mentioned earlier, such an increase in the number of crop-cutting experiments will cause tremendous problem of organization and finance. Therefore, a new technique needs to be developed which could be adopted as an annual measure for estimation of yields for small areas, and which does not necessitate heavy expenditure and is capable of being handled by the available local field agency with marginal additions to provide for requisite supervision of field work and for processing of data. With this object in view, research investigations have recently been conducted, utilising the technique of double sampling by the Institute of Agricultural Research Statistics. The yield of the chosen crop is estimated by eye appraisal from a large sample of fields growing that crop while for a sub-sample of these fields the yield is estimated physically by the method of crop cutting. The eye estimate is usually made 10-15 days before the harvest of the crop. The method and other details have already been given by Panse and others (1966). Some of the operational and theoretical implications of the procedure have been discussed in the following paragraphs.

The method of obtaining yield by eye appraisal is a subjective procedure and cannot be free from bias. If however, a good correction factor for such a bias can be estimated by using objective method of crop-cutting, the method of eye estimate which is relatively much cheaper, can be advocated for routine use for estimating crop yield. The purpose of this research investigation is not to replace the objective method of crop-cutting but to obtain auxiliary information on yield to reduce the number of crop-cutting to the minimum. The success of the investigation is ultimately dependent on the relationship between the eye estimate and crop-cutting yields.

Sampling Design

Generally, a block consists of 8-10 groups, each of 8-10 contiguous villages, known as Village Level Worker's (V.L.W.) circle. For the survey, the VLW circles were taken as strata, the village as the primary sampling unit, a field growing the crop under study within the village as the second stage unit and randomly located plot of standard dimension in the field as the ultimate sampling unit. From each stratum, 4-5 villages and from each sampled village 4 fields growing the crop were selected by simple random method for eye appraisal of yield. Thus, from each VLW circle 16-20 fields were selected for eye estimation. From these fields, a random sub-sample of four fields, by taking two villages and two fields per village was selected for crop-cutting.

The VLW, who has received agriculture-oriented training and is responsible for agricultural extension work in his jurisdiction, was asked to use his judgement to estimate the yield of sampled fields by eye appraisal by careful inspection of the crop within a fortnight of the harvest of the crop.

The crop-cutting work was entrusted with the Patwari, a village Land Record Official, who is responsible for assistance in the conduct of normal crop-cutting. Within a block there are generally 15-20 Patwaris, each incharge of 4-5 villages. The workload on each Patwari was, therefore, not more than four crop-cuttings, which he could easily do. For the VLW also, the workload of obtaining eye estimate of yield for 20 fields is considered to be light as he has not to make any special efforts for this purpose. It is presumed that during the course of his normal tour for various extension work in these villages he would be able to visit the sample fields and obtain relevant information. The field work was supervised by statistically trained supervisors specially recruited for the purpose.

Notation :

Let

K = number of VLW circles (number of strata) in a block ;

N_i = total number of villages (first stage sampling units) in i-th stratum ($i = 1, 2, \ldots K$).

n' = number of villages selected randomly in the i-th stratum for eye estimation of yield ;

n = size of sub-sample of these n' first stage units for crop-cutting.

M_{ij} = number of fields growing the crop (second stage sampling units) in the j-th village of i-th stratum ($j = 1, 2, \ldots, N_i$; $i = 1, 2, \ldots K$) ;

m' = the number of crop fields selected for eye estimate of yields in each of the sample villages ;

m = size of the sub-sample of these m' second stage sampling units for crop-cutting ;

y_{ijl} = estimated yield of the crop by crop-cutting in the l-th field in the j-th village of i-th stratum ; and

x_{ijl} = eye estimated yield of the crop in the l-th field in j-th village of i-th stratum ($l = 1, 2 \ldots m', j = 1, 2 \,.\, n'$; $i = 1, 2, \ldots K$)

For simplicity, M_{ij}, number of fields growing the crop in a village may be assumed to be large. For practical convenience it may be considered desirable to allot the same workload to each of the field workers and therefore, number of villages and number of fields sampled from each stratum have been taken to be the same. The advantage of the double sampling procedure in this particular case will depend upon how good are the eye-estimates of yield by VLW's. Although the V.L.W's are trained in agriculture and fully acquainted with the local conditions, but to start with, they have little experience in obtaining eye-appraisal of yields. In the pilot investigation it was suggested that the V.L.W., in addition to applying his own judgement, might consult experienced cultivators in the village for arriving at correct idea of yields. It is, however, expected that after a little experience, this will be the most appropriate agency for this task.

For this type of sampling design, well-known methods of ratio and regression are generally suggested.

The ratio estimate will be of the form

$$\overline{y}_R = R_x \, \bar{x}_{n'm'}$$

where

$$R = \bar{y}_{nm} / \bar{x}_{nm}$$

$$\bar{y}_{nm} = \frac{1}{n_0 m} \sum_{i=1}^{K} \sum_{j}^{n} \sum_{1}^{m} y_{ijl},$$

$$\bar{x}_{mn} = \frac{1}{n_0 m} \sum_{i=1}^{K} \sum_{j}^{n} \sum_{l}^{m} x_{ijl},$$

$$\bar{x}_{n'm'} = \frac{1}{n_0' m'} \sum_{i=1}^{K} \sum_{j}^{n'} \sum_{1}^{m'} x_{ijl},$$

$$n_0 = Kn$$

and

$$n'_0 = Kn'$$

An estimate of the variance is given by

$$\hat{V}(\bar{y}_R) = \left(\frac{1}{n_0} - \frac{1}{N_0}\right) \bar{s}_{by}^2 + \frac{1}{N_0 m} \bar{s}_{wy}^2$$

$$-2R\left[\left(\frac{1}{n_0} - \frac{1}{n_0'}\right) \bar{s}_{bxy} + \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right) \bar{s}_{wxy}\right]$$

$$+R^2\left[\left(\frac{1}{n_0} - \frac{1}{n_0'}\right) \bar{s}_{bx}^2 + \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right) \bar{s}_{wx}^2\right]$$

where

$$\bar{s}_{bxy} = \frac{1}{K(n-1)} \sum_{i=1}^{K} \sum_{j}^{n} (\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot})(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot}),$$

$$\bar{s}_{wxy} = \frac{1}{n(m-1)} \sum_{i=1}^{K} \sum_{j}^{n} \sum_{l}^{m} (x_{ijl} - \bar{x}_{ij})(y_{ijl} - \bar{y}_{ij\cdot}),$$

$$\bar{s}_{by}^2 = \frac{1}{K(n-1)} \sum_{i=1}^{K} \sum_{j}^{n} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2,$$

$$\bar{s}_{wy} = \frac{1}{n_0(m-1)} \sum_{i=1}^{K} \sum_{j}^{n} \sum_{l}^{m} (y_{ijl} - \bar{y}_{ij\cdot})^2,$$

$$\bar{y}_{ij\cdot} = \frac{1}{m} \sum_{l}^{m} y_{ijl},$$

$$\bar{y}_{i\cdot\cdot} = \frac{1}{n} \sum_{j}^{n} \bar{y}_{ij\cdot},$$

and

$\bar{x}_{ij\cdot}$ $\bar{x}_{i\cdot\cdot}$, $\bar{s}_{bx}^2$ and $\bar{s}^2_{wx}$ are the corresponding form for variate x.

A linear regression estimate can be written as

$$\bar{y}_{lr} = \bar{y}_{nm} + b(\bar{x}_{n'm'} - \bar{x}_{nm})$$

where the value of b can be shown to be

$$b = \frac{\left(\frac{1}{n_0} - \frac{1}{n_0'}\right) \bar{s}_{bxy} + \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right) \bar{s}^2_{wxy}}{\left(\frac{1}{n_0} - \frac{1}{n_0'}\right) \bar{s}_{bx}^2 + \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right) \bar{s}_{wx}^2}$$

An estimate of variance of $\bar{y}_{lr}$ can be approximated as

$$\hat{V}(\bar{y}_{lr}) = (1 - r^2) \left[\left(\frac{1}{n_0} - \frac{1}{N}\right) \bar{s}_{by}^2 + \frac{1}{N_0 m} \bar{s}_{vy}^2\right]$$

$$+ r^2 \left[\left(\frac{1}{n_0'} - \frac{1}{N_0}\right) \bar{s}_{by}^2 + \left\{\frac{1}{N_0 m} - \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right)\right\} \bar{s}_{wy}^2\right]$$

where $r^2 = \frac{q_{xy}^2}{q_{xx} q_{yy}}$

with

$$q_{xy} = \left(\frac{1}{n_0} - \frac{1}{n_0'}\right) \bar{s}_{bxy} + \frac{1}{n_0}\left(\frac{1}{m} - \frac{1}{m'}\right) \bar{s}_{wxy}$$

Expressions for q_{xx} and q_{yy} may be obtained by substituting x for y and y for x respectively in q_{xy}.

4. Discussion

Efficiency of the ratio and regression methods of estimation depends on the relationship between x and y. There are some basic assumptions that must be satisfied before recommending these methods. Both methods are based on large sample concept. In addition, certain functional relationship must be satisfied between x and y. The conditions of functional relationship for application of ratio method of estimation is simple. The functional relationship between x and y should be linear and lines should pass through the origin. If these conditions are satisfied, even for moderate sample size with moderate correlation coefficient, the ratio method of estimation can be recommended without introducing any serious bias. Regression method of estimation needs some more conditions to be satisfied, besides the relationship between x and y being linear and the population being large. One of the implicit conditions for regression approach is that residual variance of y about the regression line is constant. However, in surveys in which the relation of y and x is thought to be linear, the use of regression methods of estimation may be helpful without having to assume exact linearity or constant residual variance.

It may be worthwhile to examine how far the basic assumptions are satisfied in the application of double sampling approach for estimating the yield of crops for small area. Eye estimates cannot be considered as a fixed quantity as its value depends upon the person who makes them. However, in case of a VLW Circle only one man is made responsible to make eye estimate and therefore, x can be considered to be fixed for a sampled field. For a given x (eye estimate), y(crop-cutting result) follows certain frequency distribution and this is one of the assumptions in applying regression method of estimation. We may assume that when x is zero y will be zero and therefore, both methods of estimation, ratio as well as regression can be applied in this particular case.

As mentioned earlier both methods are subject to bias and it is necessary to examine the magnitude of the bias before recommending any of the two methods for routine adoption. Detailed expressions of bias terms for the two methods when the sampling is multi-stage have been worked out by Grewal (1966).

Three estimates, simple estimate based on crop-cutting results alone, ratio estimates and regression estimates along with their biases have been worked out for the data collected during 1965-66 in the pilot surveys for estimating the yield of wheat crop at the Community Development Block Level in the district of Patna (Bihar State). The results are given in Table I. The estimates based on crop-cutting alone have been assumed to be free from bias. It can be seen from the Table that the relative bias in case of both the estimates namely, ratio and regression is almost negligible.

In a method of this type, there is always a risk of collusion in the results obtained on the basis of eye estimates and those obtained on the basis of crop-cutting. It may happen that crop-cutting results may be influenced by eye estimates and vice-versa. There are several ways to check up the bias introduced on account of these reasons. Since two sets of fields, one where eye estimate alone is obtained and the other where along with eye estimate crop-cuttings are also conducted, are two independent sub-samples coming from the same population, the mean results based on eye estimates for these two should be comparable. The results are given in Table 1 (cols. 6-9). It can be seen that in none of the blocks the means are differing significantly. It may, therefore, be concluded that there is little fear of any collusion between eye estimates and crop-cutting results. It may, however, be emphasised that such collusions can be fully avoided only by intensifying the supervision over the field work which was done in these pilot investigations.

Since eye estimate (x) in this particular case may not be considered always free from error, the methods suggested by Wald (1940) and Bartlett (1949) have been applied to study the effect of such errors on the estimate. The results are presented in Table 2. It can be seen that the averages as obtained on the basis of usual regression method (Table 1) and those calculated on the basis of Wald's and Bartlett's methods agree closely. This further suggests that the bias, if any, in using usual regression method may be negligible.

Another statistical check was employed to study the biasedness in the double sampling procedure including that due to field work. Since random crop-cutting procedure is considered to be free from bias it was thought desirable to compare the estimate obtained through the double sampling procedure with that obtained on crop-cutting survey alone. For this purpose the study was undertaken in a few blocks selected in Dhulia (Maharashtra), Patna (Bihar) and Meerut (U.P.) where special independent crop-cutting experiments (different from those conducted under double sampling procedure) were conducted on wheat during 1965-66. Results of the study are given in Table 3. It may be observed that the average yield obtained on the basis of these independent crop-cutting experiments and those obtained from the double sampling procedure are not significantly different. This provides further confidence in the technique of double sampling for obtaining estimates for smaller areas.

Acknowledgement :

I gratefully acknowledge the help given by Shri S.S. Pillai in preparing this paper. I am also thankful to Shri J.N. Garg and Shri A.K. Srivastava who helped in collecting the material.

REFERENCES

1. Watson, D.J. (1937) : "The estimation of leaf area in field crops", Jour. of Agri. Sciences, Vol. XXVII.
2. Neyman. J. (1938) : "Contribution to the theory of sampling human populations", Jour. of Amer. Stat. Assoc. Vol. 33.
3. Snedecor, and King (1942) : "Recent developments in sampling for agricultural statistics", Jour. of Amer. Stat, Assoc. Vol. 37.
4. Yates, F. (1960) : "Sampling methods for censuses and surveys", 3rd Edition, Charles Griffin and Co. Ltd. London.
5. Cochran, W.G. (1963) : "Sampling Techniques", 2nd Edition, John Wiley and Sons, New York.
6. Sukhatme, M.B and Koshal, R.S. (1959) : "A contribution to double sampling", Jour. of Ind. Soc. Agri. Stat. Vol. XI.
7. Singh, D. and Singh B.D. (1965) : "Double sampling for stratification" on successive occasions", Jour. of Amer. Stat. Assoc. Vol. 60.
8. I.C.A.R. Publication (1951) : "Sample surveys for the estimation of yield of food crop " I.C.A.R. Bulletin Series No. 1.
9. Panse V.G., Rajagopalan, M. and Pillai, S.S. (1966) : "Estimation of crop yields for small areas". Biometrics Vol. 22, No. 2.
10. Wald A. (1940) : "The fitting of straight lines if both variables are subject to error", Ann. Math. Stat. Vol. 11.
11. Barlett, M.S. (1949) : "Fitting a straight line when both variables are subject to error". Biometrics, Vol. 5.
12. Grewal, G.K. (1966) : "Use of double sampling with special reference to block level estimates", unpublished thesis submitted for M. Sc. Degree in Agricultural Statistics at I.A.R.S., New Delhi.

TABLE 1

Only crop-cutting experiments are considered to be subject to error

State : Bihar District : Patna Crop : Wheat Year : 1965-66

Blockwise estimate of yield rate in Kilogram per hectare

Block	No. of crop-cutting expts.	No. of eye estimate	Crop cutting mean	% S.E.	Eye estimate mean from c.c. fields (Kg/Ha)	% S.E.	Eye estimate mean from other fields (Kg/Ha)	% S.E.	S.E. of diff. bet. eye est. from c c. and other fields	Ratio estimate			Regression est.				Coff. of correlation
										Est. (Kg/Ha)	% S.E.	% rel. bias	Reg. Coeff.	Est. (Kg/Ha)	% S.E.	% rel. bias	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Bihta	50	200	959	5·0	978	6·9	877	3·3	73·56	882	6·2	0·27	0·37	929	4·6	—0·21	0·52
2. Dhararua	44	152	794	8·2	806	9•5	772	6 6	92·11	726	8 6	0·23	0·55	754	7·6	—0·13	0·65
3. Maner	24	64	827	8·7	777	8·2	797	5·9	79·31	839	7·4	0·08	0·90	837	7·4	0·01	0·81
4. Paliganj	50	220	675	12·6	736	8·2	717	3·9	66·54	652	10·8	0·04	0·96	651	10·8	—0·15	0•68
5. Naubatpur	46	188	630	9·2	795	7•0	812	3 2	61·16	631	8 0	0·14	0·61	630	7·9	0·02	0·59
6. Hernaut	48	148	722	6·1	721	5·5	721	4·1	49·27	725	5·9	0·04	0·56	724	5·7	—0·31	0·52
7. Bakhtiarpur	46	148	1045	5·3	1019	5·5	1077	4·6	74·87	1122	4·7	—0·04	0·58	1088	4·5	—0·11	0·59
8. Pandarak	50	164	726	6·4	860	5·9	859	3·6	59·71	729	5·7	0·08	0·59	728	5·5	0·14	0·65
9. Mokameh	50	168	658	6.3	992	5·9	1090	3·7	71·06	707	5·1	0 02	0·46	692	5·0	0·08	0·66
10. Chandi	50	192	762	6·7	889	8·3	894	4·9	87·31	753	6·7	0·23	0·45	756	5·9	0·00	0 66
11. Barh	50	232	854	4·4	919	4·8	854	2·1	47·57	807	3·3	0·05	0·72	816	3·1	0·00	0·84
12. Hilsa	40	156	779	4·5	867	6·4	959	5·2	75·05	864	3·7	—0·07	0·50	825	3·2	—0·05	0·80
13. Asthawa	44	196	780	5·6	869	5·7	838	2·4	53·66	739	5·7	0·15	0·48	758	5·0	—0·04	0·55
14. Noorsarai	40	140	833	5·8	901	5·7	863	3·9	61·26	799	4·3	0·01	0·79	803	4·2	—0·01	0·86
15. Islampur	42	140	786	13·2	833	10·8	784	4·6	96·63	752	10·9	0·32	0·95	750	10·9	—0·58	0·83

TABLE 2

Crop cutting values and eye estimates considered to be subject to error

State : Bihar District : Patna Crop : Wheat Year : 1965-66

Block	*Wald's estimates*		*Bartlett's estimates*	
	Regression equation	*Estimate kg/Ha*	*Regression equation*	*Estimate kg/Ha*
1	*2*	*3*	*4*	*5*
1. Bihta	$Y=0\cdot26x+706\cdot6$	938	$Y=0\cdot12x+814\cdot2$	928
2. Dhararua	$Y=0\cdot61x+300\cdot2$	786	$Y=0\cdot62x+300\cdot8$	797
3. Maner	$Y=0\cdot84x+174\cdot7$	833	$Y=0\cdot78x+216\cdot9$	833
4. Paliganj	$Y=0\cdot74x+128\cdot8$	674	$Y=0\cdot91x+\ 13\cdot6$	682
5. Naubatpur	$Y=0\cdot42x+299\cdot4$	640	$Y=0\cdot28x+360\cdot8$	595
6. Harnaut	$Y=0\cdot64x+261\cdot5$	719	$Y=0\cdot55x+321\cdot0$	720
7. Bakhtiarpur	$Y=0\cdot55x+483\cdot5$	1081	$Y=0\cdot49x+544\cdot5$	1077
8. Pandarak	$Y=0\cdot69x+132\cdot2$	747	$Y=0\cdot72x+116\cdot9$	762
9. Mokameh	$Y=0\cdot52x+726\cdot1$	726	$Y=0\cdot36x+278\cdot4$	685
10. Chandi	$Y=0\cdot27x+518\cdot2$	775	$Y=0\cdot07x+661\cdot2$	728
11. Barh	$Y=0\cdot75x+167\cdot7$	824	$Y=0\cdot92x+\ 19\cdot9$	833
12. Hilsa	$Y=0\cdot54x+313\cdot6$	812	$Y=0\cdot60x+261\cdot2$	820
13. Asthawa	$Y=0\cdot44x+397\cdot5$	774	$Y=0\cdot37x+454\cdot0$	771
14. Noorsarai	$Y=0\cdot71x+194\cdot4$	817	$Y=0\cdot83x+\ 91\cdot8$	821
15. Islampur	$Y=0\cdot62x+274\cdot0$	765	$Y=0\cdot66x+229\cdot2$	763

TABLE 3

Results of special crop-cutting experiments and double sampling estimates conducted on wheat crop, 1965-66.

State	*District*	*Block*	*Special crop cutting experiments*			*Double sampling estimates*			
			No. of crop-cutting	*Av. yld. (Kg/ha)*	*% S.E.*	*No. of crop-cutting expts.*	*No. of eye-estimates*	*Regression estimates (Kg/ha)*	*% S.E.*
1. Uttar Pradesh	Meerut	Hapur	78	1275	4·36	44	184	1220	3·82
		Kharkhoda	70	1098	4·44	40	128	1182	5·45
		Gharmukteswer	66	1220	3·71	48	196	1052	6·32
		Simbhaoli	69	1194	4·00	36	140	1173	4·85
2. Maharashtra	Dhulia	Dhulia	64	678	9·11	34	140	587	6·02
		Sakri	72	707	6·11	40	164	840	7·46
		Shindkheda	36	638	10·54	34	175	529	10·43
		Nandurbar	62	552	6·31	29	139	577	10 45
		Shahada	42	612	10·23	14	104	809	14·14
3. Bihar	Patna	Naubatpur	50	654	2·79	46	186	668	7·11
		Asthawan	40	745	2·55	44	194	762	2·93
		Biharshar	42	554	4·01	48	130	563	4·95

ANALYSIS AND CONSTRUCTION OF FRACTIONAL AND CONFOUNDED FACTORIAL DESIGNS WITH EMPHASIS ON THE ASYMMETRICAL CASE

by

J.N. SRIVASTAVA*

1. Introductiion. This paper presents a summary of the work done by the author both singly and jointly, in the development of the theory of confounded and/or fractionally replicated asymmetrical factorial designs. In general, the results in the asymmetrical case have been obtained by extending the corresponding symmetrical ones.

Fisher (1925) and Yates (1937), gave us the basic concepts and developed the first batch of confounded designs. A general theory of confounding in the symmetrical case using finite fields and geometries was first started by Bose and Kishen (1940), and later presented in a complete form by Bose (1947). Similarly, a complete theory for most asymmetrical designs likely to arise in practice, was first developed by Kishen and Srivastava (1959) by generalising Bose's methods of finite geometries, and supplementing them by new ones. A more powerful technique of construction of balanced asymmetrical designs was given by Das (1960) who obtained them as fractions of symmetrical designs. The work was preceded by that of Nair and Rao (1948), Li (1944) and others, who used different approaches to the problem. The concept of a fractionally replicated factorial experiment was first introduced by Finney (1945) with important advances made by Plackett and Burman (1946), Kishen (1948), Rao (1950), and others (1, 9, 11, 14, 20, 23, 6, 18). Recently, Bose (1961) established an important connection between this whole area and coding theory.

2. Definition of effects

Consider a symmetrical factorial experiment with m factors at s levels each, or briefly $SFE(s^m)$. Writing the treatment combinations or 'assemblies' in the lexicographic order $a_1^{j_1} a_2^{j_2} \dots a_m^{j_m}$,

$$0 \leqslant j_r \leqslant s-1 \,;\; r=1, 2, \dots m.$$

In general, we shall use the same symbol both for an assembly and the 'observed yield' for it. If in the symbol for a treatment, the exponent j_r of a_r is zero, we shall omit this a_r from the symbol. If $j_r=0$ for all r, the

---

*University of Nebraska. Present address : University of Colorado.

symbol will be writtent ϕ . Exactly similar notation will be used for interaction components except that a's will be replaced by A's and ϕ by μ (the general total). It is well-known [see, *e.g.* Bose (1947)], Kempthorne (1952) that each interaction (including main effect) degree of freedom can be expressed as a linear contrast of all treatment combinations :

$$A_1^{k_1} \ A_2^{k_2} \ \ldots \ A_m^{k_m} = \sum_{j_1, \ldots, j_m} d^{j_1, j_2, \ldots, j_m}_{k_1 k_2, \ldots, k_m} \left(a_1^{j_1} \ a_2^{j_2} \ \ldots \ a_m^{j_m} \right), \qquad \ldots(2.1)$$

where $k_t = 0, 1, \ldots, s_t - 1 \, ; \, t = 1, 2, \ldots m.$

By choosing the above d coefficients suitably the interaction contrasts are expressed differently.

3. λ-operator

Let T be any fractional replicate from $SFE(s^m)$, where any assembly from $\Omega_{m,s}$ can occur 0, 1 or more times. Let us have a sub-assembly

$$\theta = a_{i_1}^{j_1} \ a_{i_2}^{j_2} \ \ldots \ a_{i_r}^{j_r} \ ; \ 1 \leqslant r \leqslant m, j\text{'s} \in (0, 1, \ldots, s-1) \, ;$$

from any r of the m factors.

We define

$\lambda(\theta, T)$ = Number of assemblies in T in which the symbol θ occurs. ...(3.1)

Let $\theta' = a_{i'_1}^{j'_1} \ a_{i'_2}^{j'_2} \ \ldots \ a_{i'_r}^{j'_r}$ be any other assembly. Consider the linear combination $(\beta\theta + \beta'\theta')$, where β and β' are real numbers.

We define

$$\lambda(\beta\theta + \beta'\theta', T) = \beta\lambda(\theta, T) + \beta'\lambda(\theta', T) \qquad .. (3.2)$$

If T_1 and T_2 are any two fractions, denote by $T_1 + T_2$ the fraction obtained by pooling T_1 and T_2 together. If $| T |$ denotes the number of elements in T, then clearly

$$| T_1 + T_2 | = | T_1 | + | T_2 | \qquad \ldots(3.3)$$

$$\lambda(\theta, T_1 + T_2) = \lambda(\theta, T_1) = \lambda(\theta, T_2) \qquad \ldots(3.4)$$

Equation (3.4) thows that the λ – operator is linear, over different fractions.

The above definitions can be immediately extended to the

$$AFE(s_1 \times s_2 \times \ldots \times s_m)$$

by keeping all of them intact except that the superscript j_t (of $\left(a_{i_t}^{j_t} \right)$ in 0) now takes the values $(0, 1, \ldots, s_{i_t} - 1)$ instead of $(0, 1, \ldots, s-1)$.

In order to illustrate the utility of the λ-operator we have discussed below some theory for $SFE(2^m)$.

Consider the ring P of polynomials in the set of symbols

$$a_r^j \ (r=1, 2, \ldots, m \ ; \ j=0, 1, 2, \ldots, s-1).$$

To this set adjoin two symbols a_0^0, a_0^1 which shall act respectively as additive and multiplicative identities. We further assume

$$(i) \ \left(a_r^j \right)^t = \left(a_r^j \right), \text{ for all integers } t \neq 0$$

$$= \left(a_0^1 \right), \text{ if } t=0. \qquad \ldots(3.5)$$

$$(ii) \ a_r^0 + a_r^1 + \ldots + a_r^{s-1} = a_0^1, \text{ for all } r. \qquad \ldots(3.6)$$

(*iii*) The usual commutative, associative and distributive laws hold.

Now let $s=2$. Any polynomial P in the above $2m+2$ symbols can then be expressed as a polynomial in the $m+2$ symbols

$$a_r^1 \ (r=0, 1, \ldots m) \text{ and } a_0^0 .$$

For example

$$a_r^1 \, a_r^0 = a_r^1 \left(a_0^1 - a_r^1 \right) = \left(a_r^1 \, a_0^1 \right) - \left(a_r^1 \right)^2 = a_r^1 - a_r^1 = a_0^0 \qquad \ldots(3.7)$$

For any fraction T, we define

$$\lambda\left(a_0^{\overline{0}}, T \right) = 0, \ \lambda\left(a_0^1, T \right) = |T| \qquad \ldots(3.8)$$

Example I

Consider the following fraction T from $AFE(2^4 \times 3^2)$ in 48 assemblies, given by $T=[T_1 \otimes T_2]+[T_3 \otimes T_4]+[T_5 \otimes T_6]$, where the arrays T_i are indicated below.

$$\begin{array}{c}T_1\\ \begin{vmatrix}0&0&0&0\\1&1&1&0\\1&1&0&1\\1&0&1&1\\0&1&1&1\end{vmatrix}\end{array}\quad \begin{array}{c}T_2\\ \begin{vmatrix}0&0\\1&2\\2&1\end{vmatrix}\end{array}\quad \begin{array}{c}T_3\\ \begin{vmatrix}1&1&1&1\\0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{vmatrix}\end{array}\quad \begin{array}{c}T_4\\ \begin{vmatrix}0&1\\1&0\\2&2\end{vmatrix}\end{array}\quad \begin{array}{c}T_5\\ \begin{vmatrix}1&1&0&0\\1&0&1&0\\1&0&0&1\\0&1&1&0\\0&1&0&1\\0&0&1&1\end{vmatrix}\end{array}$$

$$\begin{array}{c}T_6\\ \begin{vmatrix}1&1\\2&0\\0&2\end{vmatrix}\end{array} \qquad \cdots(3.9)$$

The kronecker product notation is the standard one. Thus $(T_1 \otimes T_2)$ consists of the 15 assemblies (000000), (000012), (000021), (111000), (111012)..., (011112, (011121), where the first 4 elements in any assembly refer to the factors at 2 levels (called A-factors), and the others to 3-level or B-factors. Thus (011121) stands for

$$a_1^0\ a_2^1\ a_3^1\ a_4^1\ b_1^2\ b_2^1\ .$$

Let

$$P_1 = \left(a_1^1 - a_1^0\right)\left(a_3^1 - a_3^0\right)\left(b_1^2 - b_1^0\right)\left(b_2^2 - 2b_2^1 + b_2^0\right),$$

$$P_2 = \left(a_2^1 - a_2^0\right)\left(a_3^1 - a_3^0\right)\left(b_1^2 - b_1^0\right)\left(b_2^2 - b_2^0\right),$$

Then using (3.5, 3.6),

$$P \equiv P_1 P_2 = Q_1 Q_2,$$

where $$Q_1 = \left(a_1^1 - a_1^0\right)\left(a_2^1 - a_2^0\right),$$

$$Q_2 = \left(b_1^2 + b_1^0\right)\left(b_2^2 - b_2^0\right)$$

assuming $b_j^i\ b_j^{i'} = 0 (i \neq i' = 0, 1, 2)$

$$\cdots(3.10)$$

Also, clearly

$$\lambda(Q_1Q_2,\ T_1\otimes T_2)=\lambda(Q_1,\ T_1).\ \lambda(Q_2.T_2).$$

$$\lambda(Q_1,\ T_1)=\lambda(a_1{}^1a_2{}^1,\ T_1)+\lambda(a_1{}^0a_2{}^0,\ T_1)-\lambda(a_1{}^1a_2{}^0,\ T_1)-\lambda(a_1{}^0a^1{}_2,\ T_1)$$
$$=2+1-1-1=1$$

$$\lambda(Q_2,\ T_2)=\lambda([b_1{}^2b_2{}^2+b_1{}^0b_2{}^2-b_1{}^2b_2{}^0+b_1{}^0b_2{}^0],\ T_1)$$
$$=0+0-0-1=-1.$$

Hence $\lambda(P,\ T_1\otimes T_2)=-1$. Similarly $\lambda(P,\ T_3\times T_4)=(1)\ (1)=1$, $\lambda(P,\ T_5\times T_6)=(2)\ (0)=0$, and hence $\lambda(P,\ T)=-1+1+0=0$.

4. Normal equations

4·1. General designs without blocks

Let a denote the column vector all of assemblies in the natural order :

$$\Big[\ \phi\ ;\ a_1,\ a_2,\ \ldots,\ a_m\ ;\ a_1{}^2,\ a_2{}^2,\ \ldots,\ a_m^2\ ;\ \ldots\ldots\ ;\ a_1^{s-1}\ ,\ \ldots\ ,\ a_m^{s-1}\ ;$$
$$a_1a_2, a_1a_3,\ \ldots,\ a_{m-1}\,a_m\ ;\ a_1^{s-1}\ a_2^{s-1}\ ,\ \ldots,\ a_{m-1}^{s-1}\ a_m^{s-1}\ ; \qquad \ldots(4.1)$$
$$a_1{}^2a_2,\ a_1{}^2a_3,\ \ldots,\ a_{m-1}a_m{}^2\ ;\ a_1a_2a_3\ldots\ldots;\ \ldots\ ;\ a_1^{s-1}\ \ldots\ldots\ a_m^{s-1}\ \Big]$$

Let A denote the column vector of A's in the same order. Then equations (2.1) could be written in matrix notation as :

$$A=Da \qquad \ldots(4.2)$$

where D is $(s^m\times s^m)$ and any two of its rows are orthogonal, and each element in its first row equals 1. Denote by $\delta_i{}^2$, the sum of squares of the elements in the i-th row of D ; and $\triangle^2$, the $(s^m\times s^m)$ diagonal matrix with $\delta_i{}^2$ in the cell $(i,\ i)$; $i=1,\ 2,\ ..s^m$. Then $C\equiv\triangle D$ is an orthogonal matrix.

Thus
$$a=C'\triangle A=D'\triangle^2A \qquad \ldots(4.3)$$

First we assume that no block effects are present. Let $A+-$be partitioned as $A'=(L',\ I^{0'})$, where L ($\nu\times 1$, say) of the vector of effects which are of interest to us. Usually L will consist of one of the following : (*i*) main effects, (*ii*) general mean, main effects and 2-factor interactions, (*iii*) $L=A$ itself.

Suppose now that all elements of I_0 are zero. Let T be any fraction. Let the expected values of the assemblies in T written in the form of a vector be denoted by y^*. Let y be the vector of observations corresponding to y^*. Then we have.

Theorem 4.1. (*i*) The expectation equations (under alinear fixed effects model) are

$$\text{Exp }(y)=y^*=E'\ \triangle_0{}^2L=E'p\text{ say} \qquad \ldots(4.4)$$

where E' is obtained from D' by cutting out the $(s^m - \nu)$ columns corresponding to I_0, and also by omitting (or repeating) the rows corresponding to treatment combinations omitted (or repeated) from a to get y^*, and where Δ_0^2 is Δ^2 with the last $(s^m - \nu)$ rows and columns cut out. Note that the rows of E' are arranged in such a way as to correspond to elements of y^*.

(*ii*) The normal equations for the estimate $\hat{p}$ of p are therefore

$$M\hat{p} = x. \qquad \ldots(4.5)$$

where $M = EE'$ and $x = Ey$.

(*iii*) Corresponding to a, construct an $(s^m \times 1)$ vector Z in the following way. Let 0 be any element of a. Let the element in Z corresponding to the treatment combination θ be denoted by $Z(\theta)$.

Let $Z(\theta) =$ total yield of θ (from all repetitions of θ in T).

$$\text{if } \theta \in T.$$

$$= 0, \text{ if } \theta \notin / T. \qquad \ldots(4.6)$$

Then if D_0 is the matrix obtained from D by cutting out the last $(s^m - \nu)$ row of D, we have

$$x = D_0 Z \qquad \ldots(4.7)$$

In other words, x is obtained from Z in the same way as L is obtained from a.

From the above, it is clear that the main problem in obtaining $\hat{p}$ is the calculation and inversion of M-EE'. Also writing Var $(\hat{p}) = V$, it can be checked that V is proportional to M^{-1}, which exists o if p (and hence L) is estimable.

We have now discussed the details for finding the elements in M with reference to $AFE\ (2^{m_1} \times 3^{m_2})$. A general treatment combination is denoted by $\left(a_1^{j_1} \ a_2^{j_2} \ldots a_m^{j_{m_1}} b_1^{j_1'} \ b_2^{j'_2} \ldots b_{m_2}^{j'_{m_2}} \right)$. With L $(\nu_0 \times 1)$ containing only 2-factor and lower order effects, we will have

$$\nu_0 = 1 + [m_1(m_1 + 1)/2] + 2m_2^2 + 2m_1 m_2.$$

It can be seen that corresponding to the matrix EE' in the symmetrical case, a certain $(\nu_0 \times \nu_0)$ matrix, which may be denoted by FF', occurs in the normal equations. The rows and columns of FF' correspond to the main effects and two factors interaction components. The form of FF' is exhibited below

$$FF' = \left[\frac{\Omega_1 \ \Omega_3}{\Omega'_3 \ \Omega_2} \right] \qquad \ldots(4.8)$$

where Ω_1 is of the form :

$$\begin{array}{c|ccccccc|}
 & (\mu) & (B_i) & (B_i^2) & (B_iB_j) & (B_i^2B_j) & (B_i^2B_j^2) & (A_iA_j) \\
(\mu) & M_1 & M_2 & M_3 & M_4 & M_5 & M_6 & M_7 \\
(B_i) & M'_2 & M_{11} & M_{12} & M_{13} & M_{14} & M_{15} & M_{16} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
(A_iA_j) & M_7' & M'_{16} & M'_{24} & M'_{31} & M'_{37} & M'_{42} & M_{46}
\end{array} \quad \ldots(4.9)$$

Ω_2 is of the form :

$$\begin{array}{c|ccc|}
 & (A_i) & (A_iB_j) & (A_iB_j^2) \\
(A_i) & M_{50} & M_{51} & M_{52} \\
\cdot & \cdot & \cdot & \cdot \\
(A_iB_j^2) & M'_{52} & M'_{54} & M_{55}
\end{array} \quad \ldots(4.10)$$

and finally Ω_3 can be represented as :

$$\begin{array}{c|ccc|}
 & (A_i) & (A_iB_j) & (A_iB_j^2) \\
\mu & M_8 & M_9 & M_{10} \\
(B_i) & M_{17} & M_{18} & M_{19} \\
\cdot & \cdot & \cdot & \cdot \\
(A_iA_j) & M_{47} & M_{48} & M_{49}
\end{array} \quad \ldots(4.11)$$

The whole matrix FF' has 100 submatrices. The set of interactions to which a submatrix corresponds are shown on the left of the rows and above the columns. Thus for example M_{53} corresponds to $[(A_iB_j) : (A_i\ B_j)]$ and is therefore of size $m_1m_2 \times m_1m_2$, M_{45} corresponds to $[(B_i^2B_j^2)] : (A_iB_j^2)]$ and is of size $\binom{m_1}{2} \times m_1m_2$, etc. Each element of FF' thus corresponds to a unique pair of interactions. The element in the cell corresponding to the pair of interactions (x, y) will be written $\epsilon(x, y)$. Thus for example $\epsilon(B_1^2B_2, B_3^2B_4)$ lies in M_{36}, $\epsilon(B_1B_2, B_1^2)$ in M_{21}' etc.

The calculation of FF' can be established on the above lines.

4.2. Semi-regular designs without blocks

Consider $SFE(s^m)$. A fraction $T(= T_1 + T_2 + \ldots\ldots T_f)$, where T_i $(i = 1, 2, \ldots\ldots, f)$ is the set of all assemblies lying on the flat given by

$$Ax = c_i \quad \ldots(4.12)$$

where $A(a \times m)$, and $c_t(a \times 1)$ are over $GF(s)$, and $x' = (x_1, \ldots\ldots, x_m)$ is called semi-regular. We consider a restricted treatment of the case $s = 3$ only. Here we assume for simplicity that A has the property R_3, namely that no linear combination of the rows of A has less than 3 non-zero elements. Then each T_i is an orthogonal array of strength 2. Let $C^* = [c_1, \ldots\ldots, c_f]$, and $C = C^* - AJ_{mf}'$ where J_{ab} will always denote a matrix of size $(a \times b)$ with the element 1 in each cell.

Let $\wedge = [\lambda_1, \lambda_2, \ldots\ldots, \lambda_l]$ be an $(a \times l)$ matrix, whose first column λ_1 is zero. Then $\wedge$ is called an 'alias-component' of A, if the following conditions hold : (*i*) All λ_i are distinct. (*ii*) There exist integers $j_1 < j_2$, and $j_{1i} < j_{2i}$ such that for $i = 2, \ldots \ldots, l$ the term $(\lambda_i' Ax)$ is expressible in the form $(x_{j1} + \gamma x_{j2} + \gamma_{1i} x j_{1i} + \gamma_{2i} x j_{2i})$, where γ, γ_{1i} and γ_{2i} belong to $GF(3)$, and the pair (j_1, j_2) is different from each of the pairs (j_{1i}, j_{2i}). (*iii*) There does not exist a non-null vector λ_{l+1} such that condition (*ii*) is satisfied for $i = 2, \ldots\ldots\ldots, l+1$.

Notice that γ may or may not be zero. Also since A has the property R_3 it is clear that for no i shall both γ_{1i} and γ_{2i} be zero.

If $\wedge$ and $\wedge^*$ are two alias-components of A, then $\wedge$ is called alias-equivalent to $\wedge^*$ if $\wedge^* = R \wedge P$, where R is a non-singular matrix and P is a permutation matrix. Clearly alias-equivalence is an equivalence relation, and generates equivalence classes on the set of all alias-components of A. A set $(\wedge_1, \ldots\ldots, \wedge_g)$ of alias-components of A is called minimal complete if the set has exactly one matrix belonging to each equivalence class.

The sign matrix $D_\wedge$ of an alias-component $\wedge (a \times \psi)$ is an $(l \times l)$ real diagonal matrix (diag $(d_1, \ldots\ldots, d_l)$ say) such that (*i*) $d_1 = -1$, (*ii*) for $i \geqslant 2$, $d_i = 1$, if γ_{1i} (defined above) $= 1$; otherwise $d_i = -1$.

4.3. Block Designs

Consider now a fraction T with $N = | T |$, in which the assemblies are divided into b blocks, i-th block containing the k_i assemblies T_i, where $T = T_1 + \ldots\ldots T_b$. Let $H(N \times b)$ be the incidence matrix of the design, *i.e.*, the rows of H correspond in order to the assemblies in y, and the columns to blocks, such that the element in the cell (θ, i) of H is 1, if the θ-th assembly in y is in the i-th block, and is 0 otherwise. Then $H'H = \text{diag}\ (k_1, k_2, \ldots\ldots, k_b)$; and $H'y = B$ say, is the vector of block totals. We then have

Theorem 4.2. The normal equations for estimating L are $\hat{M}^\circ \hat{L} = x^\circ$, where

$$M^\circ = (FF') - (FH)(H'H)^{-1}(H'F') \; ; \; x^\circ = Fy - (FH)(H'H)^{-1}B, \qquad \ldots(4.13)$$

where F and x are as at (4.7 and 4.8).

Theorem 4.3. Let e_1, e_2 be any two effects, and $\epsilon_{0T}(e_1, e_2)$ and $\epsilon_T(e_1, e_2)$ be the corresponding elements of M°, with and without the assumption of block effects. Also for any fraction W, let $\epsilon_W(\mu, e_j)$ denote the value of $\epsilon(\mu, e_j)$ calculated from W alone.

Then

$$\epsilon_{0T}(e_1, e_2) = \epsilon_T(e_1, e_2) - \sum_{i=1}^{b} \frac{1}{k_i} [\epsilon_{T_i}(\mu, e_1)][\epsilon_{T_i}(\mu, e_2)]. \qquad \ldots(4.14)$$

An effect e is said to be unconfounded if $\epsilon_i(\mu, e) = 0$, for all i. Clearly, if at least one of e_1 and e_2 is unconfounded, then we have $\epsilon_0(e_1, e_2) = \epsilon(e_1, e_2)$. Similarly if $\epsilon_0(\mu, e_1)$ and $\epsilon_\theta(\mu, e_2)$ are respectively equal to the contents d_{i1} and d_{i2}, for $\theta \in T_i$ and for all i, then (4.14) implies, that $\epsilon_{0T}(e_1, e_2) = 0$.

4.4. Optimality Criteria

We *shall now consider different optimality criteria for selecting* a fraction T. A basic criterion (G_1 say) for screening undesirable fractions in most situations is that the selected T should be economic, *i.e.* $|T|$ should be desirably small. Also, N_e, the available number of degrees of freedom for error, should not be too large or small. In this paper we generally take $5 \leqslant N_e \leqslant 30$. A fraction T is called orthogonal, if $(V)_T$ is a diagonal matrix. However, the value of N (and hence in general N_e) is generally too large for orthogonal fractions.

The other criteria deal with the properties of $(V)_T$. Let T_1 and T_2 be two competing fractions. Then T_1 is said to be 'better than' T_2 according to (G_2) the largest root criterion, if $\underset{\max}{\mathrm{ch}}(V)_{T_1} < \underset{\max}{\mathrm{ch}}(V)_{T_2}$, ($G_3$) the trace criterion, $\mathrm{tr}(V)_{T_1} < \mathrm{tr}(V)_{T_2}$, ($G_4$) the generalised variance or determinant criterion, if $|(V)_{T_2}| < |(V)_{T_2}|$. For any T, the three criteria have the following physical interpretation : (G_2) $\underset{\max}{\mathrm{ch}}(V) = \sup_b[\mathrm{var}(b'\hat{L})]$. Thus choosing T so as to minimise $\underset{\max}{\mathrm{ch}}(V)_T$ means taking that T for which the maximum possible variance of any linear parametric function is a minimum. In this sense, the largest root is a minimax criterion.

(G_3) $\mathrm{tr}(V) = \int_{b'b=1} [\mathrm{var}(b'\hat{L})]db$. That is, $\mathrm{tr}(V)$ is proportional to the average of the variance of all normalised linear parametric functions. Thus in a sense, the trace criterion refers to the average variance.

(G_4)$|(V)|$ is proportional to the volume of the ellipsoid of concentration. Thus, in a sense, it refers to the volume of the region within which the true parametric point may lie with a certain probability.

5. Balanced designs

5.1. Conditions for balance

A fractional factorial design T is called completely balanced if V(or equivalently M) is symmetric with respect to the m factors, *i.e.*, V should be invariant under any permutation of the factors. For example, for SFE(2^m), if $(A_{r_1}, \ldots, A_{r_m})$ is any permutation of $(A_1, \ldots, A_m)$, then the new L is

$$L^* = (\mu \,;\, A_{r_1}, \ldots, A_{r_m} \,;\, A_{r_1 r_2}, \ldots, A_{r_1 r_m}, A_{r_2 r_3}, \ldots, A_{r_{m-1} \cdot r_m}).$$

Let the rows and columns of V be permuted so as to correspond to L^* and let V^* be the matrix so obtained. Then the requirement for complete balance is that $V = V^*$. Using λ-operator, and the results of the last section, we can prove.

Theorem 5.1. Consider AFE($s_1 \times \ldots \times s_m$). Supose L consists of general mean, main effects and first order interactions. A necessary and sufficient condition that T is balanced, assuming no block effects, is that T be a partially balanced array of strength four. We can study the structure of $(M)_T$ for a balanced T by using the concept of multidimensional partially balanced association (MDPBA) scheme, and the related algebras as given in (8)

5.2. Applications to problems on the existence of arrays.

Since F is real, FF' is clearly non-negative definite. Thus for any T(Balanced or not), each root of $(FF')_T$ is necessarily non-negative, or equivalently

$$\underset{\min}{ch}\ (FF')_T \geqslant 0 \qquad \ldots(5.1)$$

Now suppose that the question of existence of a fraction T with a given set of parameters is under consideration. Suppose further that the matrix FF' depends on T only through this set of parameters. Then we have the followin method of testing for the existence of T. From the given parameters, calculate $(FF')_T$. Then if (5.1) is notsatisfied, the fraction (or array)T doesn ot exist. Of course, if (5.1) is satisfied, it does not necessarily mean that T exists.

Notce that the above method is non-trivial only if the set of parameters does not specify F completely. However the latter is true in almost all non-trivial existence problems, and hence the above method is valuable and gives rise to an important family of new results. In particular, a large set of rather stringent necessary conditions for the existence of PB arrays of strength four (or less) with 2 or 3 symbols has been obtained in (25). These conditions however need too much space for presentation. We therefore give a rather simple example.

Consider a *PB* array T of strength four with two symbols 0 and 1, and m constraints. Let its parameters be $\mu_0 = \lambda^{0000}$, $\mu_1 = \lambda^{1000}$, $\mu_2 = \lambda^{1100}$,

$\mu_3 = \lambda^{1110}$ and $\mu_4 = \lambda^{1111}$. Furthermore, for simplicity, we suppose T to be [1, 0] symmetric with respect to triplets (Since T is of strength four, this is equivalent to baving $\mu_0 = \mu_4$, $\mu_1 = \mu_3$). We can show that (5.1) implies **Theorem 5.2.** A set of necessary conditions that T exists are

$$\begin{aligned}
&(i) \quad (m-2)\mu_1 \geqslant (m-4)\mu_2 \\
&(ii) \quad (m^2-m+2)\mu_0 - 4(m^2-5m+2)\mu_1 + (3m^2-19m+38)\mu_2 \geqslant 0 \\
&(iii) \quad \mu_0 + 4\mu_1 + 3\mu_2)[(m^2-m)\mu_0 - 4(m-1)(m-4)\mu_1 \\
&\qquad\qquad + (3m^2-19m+32)\mu_2] \geqslant m(m-1)(\mu_0-\mu_2)^2 \qquad \ldots(5.2)
\end{aligned}$$

As mentioned earlier,when T is balanced (*i.e.*, is a PBA of strength four), the roots of $(FF')_T$ can be explicitly obtained as functions of m, by using the properties of the linear associative algebras of the *F.A.* scheme.

6.1. Confounded fractionally replicated designs for SFE(2^m)

A class of 'good' designs with a desirable value of N, for which (relatively speaking) the correlations are not too high and trace (V) is low is given in (28). In fact, these seem to be the best known designs with such properties.

6.2. General methods of construction for asymmetrical designs

We now present a number of useful methods of obtaining fractional and/or confounded designs, suitable specially for the AFE's.

(I) The method of associated vectors and truncated geometries : Methods I, II and III.

Consider AFE($s_1 \times s_2 \ldots \times s_n$). We first define the basic terminology. Let t be a prime number or a prime power. Then (i) any ordered set of n elements in GF(t) will be called an n-vector in GF(t). (ii) Corresponding to a factor A at k levels, the vector $(0, 1, 2, \ldots k-1)$ in the real field will be called the level vector of A. (iii) Corresponding to the level vector of A, there is defined an 'Associated Vector' $(\beta_0, \beta_1, \beta_2, \ldots, \beta_{k-1})$ of A, where β's are elements of $GF(t)$, not necessarily distinct.

Let the total set of possible assemblies be denoted by Ω. Let us define, corresponding to each factor, an associated vector over $GF(t)$. Let $w \in \Omega$. We can think of w as an m-vector over the real field which gives us some particular level of each of the m factors. Suppose that a certain factor A is at level l in w. Suppose l^* is the corresponding element in the associated vector of A. Let the set of all factors be denoted by F. Then we construct a vector w^* (called a transformed assembly) from the vector w by replacing, for each $A \in F$, the level (l of A in w by the symbol l^*. Then w^* is an m-vector in $GF(t)$. Corresponding to the different assemblies in Ω, we get obviously $s_1 s_2 \ldots s_m = \omega$, say, vectors w^*. The set of all ω vectors w^* may be denoted by Ω^*. We now give two definitions.

Any vector in $GF(t)$ used to generate a fraction will be called a generator. If we have m factors in all, then the generator will be an m-vector. The inner product of two vectors $\alpha=(\alpha_1, \alpha_2,\ldots,\alpha_m)$ and $\gamma=(\gamma_1\ \gamma_2,\ldots, \gamma_m)$ is $\sum_{r=1}^{m} \alpha_t\, \gamma_r$, and will be denoted by $\alpha.\gamma$.

Method I of construction of a fraction may then be described as follows: For each-factor $A \in F$, define suitable associated vectors V_A in $GF(t)$, where t is a suitable prime power. Get the set Ω^* of all transformed assemblies w^*. Choose a suitable generator y and a constant c, such that $c \in GF(t)$. Obtain the set of all $x^* \in \Omega^*$, such that $y.\, x^*=c$. Let X^* be the set of all such vectors or transformed assemblies x^*. Each x^* in X^* corresponds to one or more assemblies x. Let X be the set of all assemblies which correspond to a vector in X^*. Then X is the fraction produced by Method I, and will sometimes be denoted as $X(\,t;\ V_A\,,\ A \in F;\ y,\ c)$.

The properties of the fraction will depend on the quantities given in the brackets. The method is obviously very flexible and covers a large number of cases. It covers the theory of obtaining fraction of $SFE(s^m)$ by taking $(m-k)$-flats in $EG(m,\ s)$. A study of the general properties of the method will not be made here, but we describe for illustrative purposes a $1/s_1$th fraction of $AFE(s_1{}^2 \times s_2)$ where s_1 is prime power.

We apply method I, with $t=s_1$. For factors with s_1 levels the associated vector consists of all the s_1 elements of $GF(s_1)$ in some suitable order, and for the factor at s_2 levels, this vector has s_2 distinct elements in $GF(s_2)$, again in some suitable order. If α_1 is the unit element of $GF(s_1)$, the 3-vector $(\alpha_1,\ x_1,\ \alpha_1)$ can be taken as a generator. In the fraction obtained, the main effects of the two factors at s_1 levels will be correlated.

Method II is the method of cutting out. Suppose we have an $AFE(s_1 \times s_2 \times \ldots \times s_k)$ where $s_1 \geqslant \max\ (s_2,\ldots,\ s_k)$ and is a prime power. Then we first take a $1/s_1{}^r$th fraction $(r < k)$ of the $SFE(s_1{}^k)$,by the standard methods of finite geometries. Finally this method says that we cut out all assemblies in our fraction which have any of the (s_1-s_i) given levels of the i-th factor, $i=2,\ldots,\ k$. Thus, for example, a fraction having 6 assemblies from the $AFE(5 \times 3 \times 2)$ may be obtained by first taking 1/5th fraction of the $5 \times 5 \times 5$ factorial, and then cutting out two levels say 3, 4 of the second factor, and three levels say 2, 3, 4 of the third factor. The fraction on which cutting out is done may be obtained by taking the 25 assemblies which lie on the flat $x_1+x_2+x_3=0$, in $EG(3,\ 5)$.

The method of cutting out can be shown to be a particular case of Method I, but has been described separately because of its usefulness.

Method III is a generalisation of Method I. Suppose for the $AFE(s_1 \times s_2 \times \ldots \times s_m)$, we construct a fraction $X_1(t_1\ ;\ V^1{}_A\,,\ A \in F\ ;\ y_1,\ c_1)$ by the use of Method I. We can further reduce the size of the fraction by taking (*i*) a number of t_2 which should be a prime power, (*ii*) associated vectors $V^2{}_A$ for each $A \in F$, the element of the associated vectors being members of $GF(t_2)$, (*iii*) an m-vector y_2 in $GF(t_2)$, and (*iv*) a constant e_2 in $GF(t_2)$, and then proceeding as in Method I. For this purpose, each assembly $\theta \in X$ will be converted into a transformed assemaly θ^* by using the associated vectors $V^2{}_A$. Then each θ^* is an m-vector in $GF(t_2)$. Denote the set of all treatments by θ^*. Then our new fraction denoted by X_2 $(t_2 ; V^2{}_A , A \in F ; y_2 ; c_2 ; X_1)$ consists of all assemblies z such that (*i*) $z \in X_1$, (*ii*) if z^* is the m-vector representing the transformed assembly of z, then $z^*.y_2 = c_2$. The above procedure may be repeated if we want to cut out some assemblies from X_2 also.

It can be shown that if $t_1 = t_2$, then the fraction X_2 can also be obtained directly by Method I by taking $t = t_1^2$, and using suitable associated vectors and generators.

Method III is chiefly useful in those cases where $t_1 \neq t_2$, so that different Galois fields are to be used. The situations where method III is very useful arise when the s_i are not necessarily prime powers, as for example in $6 \times 4 \times 3$, $6 \times 4 \times 4$, $6 \times 3 \times 3 \times 2$, $10 \times 5 \times 3 \times 2$, etc.

(2) **Use of various types of arrays: Methods IV and V**: Method IV will be illustrated with reference to $AFE(2^m \times 3^n)$.

Consider the $SFE(3^{m+n})$. We take the maximum number k of linear equations in $EG(m+n, 3)$ such that the assemblies lying on the flat represented by them form an orthogonal array of strength d. Let these equations be :

$$d_{i1}x_1 + d_{i2}x_2 + \ldots + d_{im}x_m + d_{i,\ m+1}x_{m+1} + \cdots + d_{i,\ m+n}x_{m+n} = g_i .$$

We first consider the case $d = 4$. To obtain a fraction for the $2^m \times 3^n$ factorial, we replace the level 2 of each of the m-factors $X_1, X_2, \ldots, X_m$ by the symbol 1, *i.e.* we make a transformation on the levels represented by (0: 0, 1 : 1, 2 : 1). The set of assemblies resulting from this transformation (call them T_{41}) may be taken as a fraction.

The fraction T_{41} obviously is not symmetrical with respect to the two levels 0 and 1 of the m factors $X_1, \ldots, X_m$. To achieve symmetry with respect to these two levels we may take another fraction T_{42}. This fraction is obtained from θ_4 by making the transformation (0: 1, 1 : 0, 2: 0) on the levels of the m factors $X_1, X_2 \ldots . X_m$. Then for our design, we take $2 \times 3^{m+n-k}$ assemblies given by $T_4 = T_{41} + T_{42}$.

To reduce the size of the fraction below that obtainable above, we take $d < 4$, and pool several such arrays. Good asymmetrical factorial fractions can be obtained by starting with arrays of strength 3 or 2, and making suitable transformations of factor levels.

Method *V* is a variation of the above and is applicable to *AFE* $(2^{2m_1} \times 3^{m_2})$. Here the *A*-factors are paired together, and we start with an array with 3 or 4 symbols and apply transformations as in an Method IV.

Design obtained by the above methods will be found elswehere particularly in [26]. In [29], plans for estimating main effects when two factor interactions are not negligible are discussed for $2^m \times 3^n$ designs.

Several methods of construction of confounded asymmetrical factorial designs have been discussed in (17).

REFERENCES

1. Addelman, Sidney. Symmetrical and asymmetrical fractional factorial plans, Technometrics, 1962, 4, 47-58.
2. Bose, R.C. Mathematical theory of the symmetrical factorial design, *Sankhya*, 1947, 8, 107-166.
3. Bose, R.C. On some connections between the design of experiments and information theory. Bull. Int. Stat. Inst., 1961, 38, 257-271.
4. Bose, R.C. and Kishen, K. On the problems of confounding in the general symmetrical factorial design. *Sankhya*, 1940, 5, 21-36.
5. Bose, R.C. and Mesner, D.M. On linear associative algebras corresponding to the association schemes of partially balanced designs. Ann. Math. Stat., 1959, 30, 21-38.
6. Bose, R.C. and Srivastava, J.N. Mathematical theory of the factorial designs. I. Analysis, II. Construction. Bull. Int. Stat. Inst., 1964, 40(2*e*), 780-794.
7. Bose. R.C. and Srivastava, J.N. Analysis of irregular factorial fractions. *Sankhya*, Series A, 1964, 26, 117-144.
8. Bose, R.C. and Srivastava, J.N. Multidimensional partially balanced designs and analysis with applications to partially balanced factorial fractions. *Sankhya*, Series A, 1964, 145-168.
9. Chakravarti, I.M. Fractional replications in asymmetrical factorial designs and partially balanced arrays. *Sankhya*, 1956, 17, 143-164.
10. Connor W.S. and Young, Shirley. Fractional factorial designs for experiments with factors at two and three levels. Applied Mathematical Series, National Bureau of Standards, U.S. Govt. Printing Office, Washington, D.C., 1959.
11. Das, M.N. Fractional replicates as asymmetrical factional designs. Jour. Ind. Soc. Agri. Stat., 1960, 12, 159-174.
12. Finney, D.J. The fractional replication of factorial experiments. Ann. Eugen., 1945, 12, 291-301.
13. Fisher, R.A. The Design of Experiments, 7th edition, Hafner Publishing Co., New York, 1960. (Ist ed., 1925).
14. Kempthorne, Oscar. The Design and Analysis of Experiments John Wiley and Sons, New York, 1952.
15. Kishen, K. On fractional replication of the general symmetrical factorial design. J. Ind. Soc. Agric. Stat., 1948, 1, 91-106.
16. Kishen K, and Srivastava, J.N. Confounding in asymmetrical factorial designs in relation to finite geometries, Curr. Sci., 1959, 28, 98-100.

17. Kishen, K. and Srivastava' J.N. Mathematical theory of confounding in symmetrical and asymmetrical factorial designs. J. Ind. Soc. Agric. Stat., 1960, 11, 73-110.

18. Li, Jerome C.R. Design and statistical analysis of some confounded factorial experiments. Research Bulletin 333, 1944, Iowa State College of Agriculture.

19. Nair, K.R. and Rao, C.R. Confounding in asymmetrical factorial experiments. J.R.S.S., 1948, 10, B 109-131.

20. Patel. M.S. Investigations in factorial designs. Ph. D. Thesis, University of North Carolina, Dept. of Stat., 1961.

21. Plackett, R.L. and Burman, J.P. The design of optimum multifactorial experiments. Biometrika, 1946, 33, 305-325.

22. Rao, C.R. The Theory of fractional replications in factorial experiments. *Sankhya* 1950, 10, 81-86.

23. Srivastava, J.N. Contributions to the construction and analysis of designs. University of North Carolina, Inst. of Statistics Mimeograph series No. 301.

24. Srivastava, J.N. Optimal balanced 2^n fractional factorial designs. To be included in S.N. Roy Memorial Volume, Univ. of North Carolina.

25. Srivastava, J.N. Some necessary conditions for the existence of partially balanced arrays (Abstract). Ann. Math. Stat., 1965, 36, 1079.

26. Srivastava. J.N. Tables of Optimal Symmetrical and Asymmetrical Factorial Designs (under preparation) To be published by the University of North Carolina.

27. Srivastava, J.N. Theory of Optimal nonsingular semiregular fractional factional designs. Proc. 35th Session Int. Stat. Inst., Belgrade 1965.

28. Srivastava, J.N. and Bose, R.C. Some economic partially balanced 2^n factorial fractions. To be published in Ann. Inst. Stat. Math., 1966.

29. Srivastava, J.N. and Anderson, D.A. Optimal balanced plans for estimating main effects in the presence of odd ordered interactions. (Abstract) Ann. Math, Stat., 1965, 36, 1607.

30 Yates, F. The design and analysis of factorial experiments. Imperial Bureau of Soil Science, Technical Communication No. 35, 1937.

TOWARDS A SCIENTIFIC AGRICULTURAL PRICE POLICY

by

S. C. CHAUDHRI

An important feature of the agricultural price policy in India has been the fixation by the state of certain categories of prices for some commodities. For sugarcane, minimum prices applicable to deliveries made by the farmers to the sugar mills have been fixed for over three decades now ; for cotton, floor and ceiling prices within which the trade is allowed to transact business freely have been fixed for over two decades ; for foodgrains procured by the State, procurement prices have been fixed over practically the same period ; minimum support prices are also being fixed for foodgrains since 1964-65 ; and for jute, minimum support prices have been in vogue in the last five years. The object of fixing the floor or minimum support prices has been to provide protection to the farmer against fall in market prices to unduly low levels consequent on sizeable increase in production, and thus provide him with an incentive to step up investment on land. The search for a scientific basis for fixing these prices has naturally posed several problems. In this context, the need for collection of data on cost of production has been increasingly felt.

Studies into cost of production of agricultural commodities were, in recent years, first undertaken in 1952-53 when under the guidance of Dr. V.G. Panse, the Institute of Agricultural Research Statistics (then known as the Statistical Wing of the Indian Council of Agricultural Research) undertook a pilot scheme for the estimation of the cost of production of cotton and rotation crops in Akola district. This was a methodological study designed to develop a random sampling survey techniqne for estimating the cost of raising a particular crop in physical as well as monetary terms. The pilot study was followed in 1960-61 by a three-year scheme launched by the Institute for estimating the cost of production of cotton and rotation crops in principal cotton tracts in India. Earlier, in 1955-56, a three-year project for assessing the cost of cultivation of sugarcane was initiated by the then Indian Central Sugarcane Committee. Beginning with 1954-55, the Directorate of Economics and Statistics in the Ministry of Food and Agriculture has taken up a series of farm management studies. So far, these studies have covered 22 regions in 15 States and provided data on the farm economy and input-output relationships in typical soil-climate-crop complexes.

There is at least one major difference between the approach of the IARS in the farm cost studies and that of the Directorate in the farm management studies. While the former proceeded from the point of view of the crop whose production cost was estimated, the latter considered the farm as a whole. That is while the former studies were enterprise oriented the latter were farmer-oriented. This obviously had some implications for the two sampling designs : the basis for the construction of size-groups of holdings in the IARS studies was the cumulative area under the holdings, while under the Directorate's studies it was the number of holdings. In the most recent studies of the Directorate, however, the size-groups have been determined on the basis of the cumulative acreage under the holdings. There is also some difference in the imputation procedures adopted in the two sets of studies.

With the setting up of the Agricultural Prices Commission in January 1965, with the objective of advising the Government on agricultural price policy and price structure of different agricultural commodities, the importance has grown considerably of cost data and other studies that would impart a scientific basis to price fixation. In this connection it would be useful to examine the salient features of the price determination procedures being adopted in certain other countries.

In Japan, price policy coupled with other measures has brought about an agricultural revolution. The revolution began in the Meiji era—when emphasis was laid on green and animal manuring, minor irrigation, contour bunding, etc., the like of which was being done in India too, until recently. It gained momentum in the post-war period when chemical fertilizers, improved seeds and machinery gave a tremendous encouragement to farm productivity. The most striking feature of the Meiji era was the creation of a class of forward-looking farmers, through a high land tax and public procurement of agricultural commodities at incentive prices. These two measures were intended to push the inefficient farmers out of the business of cultivation. The land tax was at times as high as 70 per cent of the value of output so that, in order to continue to earn a living wage, the farmer had perforce to improve his farming efficiency. The incentive procurement system consisted of sales by the farmer of a specified portion of his produce to the State at predetermined prices. If the farmer enhanced his farm output and sold to the Government more than the quantity specified in the original contract, he got a premium over the fixed price ; but if he could not produce enough to supply the stipulated quantity, he was penalised by a lower price. Price policy was thus used for augmenting agricultural productivity. In India, it is a moot point whether such a price policy, providing for the carrot as well as the stick, would be acceptable to the country as an instrument for achieving output goals. However, in Japan, the prices fixed by the State aim at raising not only the agricultural productivity but also the farm incomes in line with growth in other sectors of the economy. Thus, the evaluation of farm family labour in the formula for price determination is done by taking into account the rate of change in the industrial wages and family labour hours, while for hired labour it varies with the changes in both the unit hired labour cost and the amount of hired labour input. The underlying idea in this system is to ensure a certain reasonable relationship of incomes from agriculture with wages of the urban worker. This is a concomitant of that stage of overall economic development where the agriculturist is to be induced to continue in farming.

In the United States of America, price support programmes are implemented in respect of a number of agricultural commodities including foodgrains, fruits, vegetables, oilseeds, tobacco, cotton and sugarbeet. The technique consists largely of a system of commodity loans under which the farmers put their produce in a warehouse against a negotiable receipt and grade certificate. They then pledge these two documents as collateral security for a loan, the value of the produce being determined on the basis of the previously announced price for the particular quality in the concerned location. These loans are non-recourse ; that is to say, the farmer forfeits the collateral without regard to its current market value. Obviously, this system calls for a large number of public warehouses, which are in an infantile stage in India at present, and also presupposes the existence of a surplus economy. However, apart from commodity loans, price support operations in the U.S.A. are also conducted through market interventions by the Government at the farm, village or market level. These operations take the form of direct Government purchases from the farmers, announcement of Government's readiness, to purchase at the support prices whatever quantities are offered, and purchase agreements with individual farmers where the support price is guaranteed.

The support prices in U.S.A. are based on parity concept. They are determined on a formula basis with the objective of giving agricultural commodities the same purchasing power which they had in a selected base period when the prices paid and received by the farmers were in good balance. A basic price is announced for a standard grade of the commodity three to four months in advance of its sowing season. On this basic price, adjustments are made again on a formula basic for the type and grade of the commodity and for the transportation costs.

In Ceylon also price policy has had an important place in programmes of agricultural development, As, however, no information was available on farm costs in the country, it was considered that the price of imported rice should be taken as indicative of the level at which the local producer should be guaranteed the price for his produce. Some such demands are often made in our country also. But, international prices are known to fluctuate because of a number of external factors which may not necessarily be related to the internal economy of the country, and at times create embarrassing situations. Thus, in Ceylon the guaranteed producer price of paddy had to be raised from Rs. 9 per bushel in 1951 to Rs. 12 in 1952, in keeping with a similar rise in world prices. But the reverse has not been found to be practicable : after 1952, the international price came down and was quoted around

Rs. 6·50 per bushel from 1960 onwards, but it has not been possible for the Government to reduce the guaranteed price which has remained constant at Rs. 12. While the continuance of high purchase prices has enabled the Government to collect under the Guaranteed Price Scheme annually as much as 60 per cent of the paddy produced, the burden of consumer subsidy on the State exchequer has been rather large.

In Australia, the price of wheat paid to the farmer lies between the export price of wheat and the internal consumer price. This procedure is made possible by the Australian Wheat Board having complete control over the entire wheat production in the country. Australia is surplus in wheat and is a principal exporter of that commodity. Therefore, depending on the price realized from exports, it can adjust the domestic prices to the producer and consumer. This situation is clearly not applicable to India.

In the patterns of price administration mentioned above, State is the monopoly or near-monopoly buyer which is not yet the case with India. Nor, in the present state of her agricultural development, does it seem advisable for her to copy any of the systems of price determination adopted in these countries. What seems desirable at the moment is to organise studies which will reveal the true structure of our farm economy from time to time and provide guidelines for a production-oriented price policy.

Agriculture in India is in a state of transition. Farms with traditional techniques exist side by side with farms where modern technological innovations have been introduced. The input-output relationships in these two types of farming are not uniform. Therefore, to devise a suitable price policy over the years, it is necessary to gauge the rate of transformation from traditional to modern agriculture. This requires data on the farm structures and costs for the two categories of farms. With a view to collecting data on the input-output structure of farms where improved farm technology has been adopted, the Agricultural Prices Commission has initiated a Farm Record Project. Under the Project, the Agricultural Extension Officers in the 16 I.A.D.P. districts are required to maintain at least one farm record book each in respect of the selected farmers in their jurisdiction. The farmers in the I.A.D.P. areas are expected to be more innovating and to have a higher level of farm technology than elsewhere. The farm record book will elicit details of all assets, farm operations, inputs and outputs both in value and physical terms.

One aspect of Indian agriculture that requires further study is the supply response to relative changes in prices. That is, how far do shifts in acreage and productivity oecur in response to observed variations in prices ? In an economy where the production of all crops has to be increased, it is more important to improve the per acre yields. Research studies on the contribution of prices to productivity in areas where agriculture is still tradition-bound and where it has been relatively modernised should prove to be rewarding.

Another aspect is the inter-relationship between the size of farm, cost of production and efficiency of resource use. Here size can be defined by reference to a number of characteristics and so too efficiency. Optimal size could then be defined with reference to a given concept of efficiency.

A corollary to the above would be the production function analysis, for different categories of farms, so as to focalise the information which would provide the basis for securing structural changes in production methods. Such analysis should furnish information on the nature and extent of incentives required for attaining a desired level of production, or on production possibillties given the cost-price constraints.

Cost of production data are collected as relating to certain specific periods of reference. To be able to utilise such data in current price policy, they have to be brought up to date. Provided that a long time has not elapsed between the time of the collection of such cost data and their use in policy making, it should be possible to make them up to date by compiling, say, the indices of input costs. Some such indices are already being compiled by several institutions, but the method of constructing them needs to be refined.

Equally necessary are the studies of price-spread. Analysis of the price spectrum enables an assessment of the real return to the producer as also the directions and patterns in which the price realised from the consumer is distributed. Such a detailed study of prices, and various margins which it comprises, enables evolution of a price policy which scientifically takes into account the interests of both the producer and the consumer.

The farmer's terms of trade provide an insight into the overall relationship between the product prices realised by him and the prices paid by him for expenses on cultivation and on living. Parity indices would therefore unfold a picture of the changing profitability from agriculture. Factor costs affect product costs and product prices affect factor prices, which in turn affect factor costs. Thus, continuing studies of the impact of output prices on cost of cultivation and cost of living, and hence of the increase in income in real terms to the producer are necessary. Many are thus the investigations other than estimating the cost of production which have to be undertaken to evolve a scientific and realistic price policy.

RURAL LABOUR—THE TASK BEFORE THE NATIONAL COMMISSION ON LABOUR

By

B. N. DATAR*

Member-Secretary,
National Commission on Labour

The National Commission on Labour was set up in December, 1966. One of its terms of reference is to study and report in particular on measures for improving the conditions of rural labour and other categories of unorganised labour. The Commission in its broader task is required to review the changes in the conditions of labour since Independence and to report on existing conditions of labour. In drawing attention to rural labour in the terms of reference of the Commission, the then Labour Minister, Shri Jagjivan Ram in his inaugural address to the Commission on 18th January, 1967 observed :

> "Our labour policy has hitherto somehow overlooked this mass of workers even though they constitute the bulk of those who produce goods and provide services. There are laws for protecting labour in factories, mines, plantations, road transport and bidi industry. I hope that contract labour will also soon enjoy a measure of legislative protection. But apart from ineffectively or imperfectly implemented minimum wages law, the workers employed in agriculture and other rural industries have been, by and large, kept beyond the purview of labour legislation."

In responding to the inaugural address, the Chairman of the Commission, Dr. P.B. Gajendragadkar, observed that 'so far, the inarticulate and unorganised labour in rural areas and other unorganised industries did not receive the attention they deserved'. Article 43 of the Constitution seeks to establish in this country social and economic justice to all its citizens. It enjoins on the Government an obligation to ensure to all workers, agricultural, industrial or otherwise, not only employment but also a living wage and proper conditions of work. It is, therefore, in the fitness of things that the Commission should be asked specifically to study and report on measures for improving the conditions of rural labour and other categories of unorganised labour.

---

* The views expressed are the author's personal views.

How does the Commission propose to tackle this task ? It must be observed at the outset that any inquiry which has to lead to a consistent framework of policy in regard to rural labour must have a reasonable foundation of facts. At the same time, the problem of collecting facts on rural labour could be so vast that an attempt to undertake fresh inquiries would, indeed, take the Commission much beyond the time limit it has imposed on itself for the completion of its work. This time limit, it hopes to complete its work by the end of 1968, has been fixed taking into account considerations which, on the one hand, require an adequate inquiry and on the other an expeditious completion of work, in order to meet the expectations aroused about it in the public mind. It is this time constraint which decides the nature of inquiry to be undertaken in every area of the Commission's terms of reference.

The extent of the problem in terms of numbers has to be determined first. According to the memorandum prepared by the Ministry of Labour and Employment, "A vast majority of the country's working population is engaged in agriculture and allied occupations. Of the total working force of about 188 million according to the 1961 Census, 133 million or roughly 70% are employed in these primary sector activities. Agricultural labour, *i.e.*, persons whose sustenance is by work on other peoples' land or livestock, accounted for about $\frac{1}{4}$th of those engaged in agriculture. Agricultural labour come from the families of marginal cultivators, backward communities and landless classes". Thus, in terms of numbers it will constitute a significant percentage of the total working force ; even assuming that the percentage goes down in the years to come, the absolute numbers are expected to be even larger according to some calculations made in the Third Plan and confirmed in the draft outline of the Fourth Plan.

To understand the problems of rural labour, one has to contend with the fact that the protection which has been given to the labour through legislation, has been admittedly ineffective. The reasons for it are many and varied. But the fact that stands out is the dispersed character of the beneficiaries of legislation and difficulties attendant on any machinery which can be envisaged for implementing the legislative provisions. The First Agricultural Labour Enquiry which was undertaken in 1950-51 revealed that the low level of living of an agricultural labour household was due in a large measure to inadequate work opportunities it usually has during a year. Any plan for improving the conditions of rural labour requires to be built up on the basis of creating additional work opportunities in rural areas. The nature of work in the countryside is such that the inquiry which has to be undertaken for a fair understanding of the problem has to be spanned over the whole year. This alone will bring out the extent of the total work opportunities needed and their distribution over the whole year. That being the nature of the difficulty and the finances required for the inquiry being large in relation to the use to which it could be put the Commission decided not to embark on special inquiries of its own covering the whole countryside ; this has been its approach in all other areas of investigation. Instead, it feels convinced that it would be adequate if the information which is already available through inquiries undertaken in the last 20 years could be subjected to careful analysis and conclusions drawn therefrom could form the basis of the Commission's recommendations. Some marginal new investigations consistent with the nature of the problem to be studied, of course, were not ruled out.

In the period since Independence two enquiries on agricultural labour have already been reported upon. The second enquiry (1956-57) which was expected to throw light on changes in conditions of labour since the period of the first (1950-51) was found somewhat inadequate for the purpose because of the change in the basis of the two enquiries. Taken together, therefore, they have raised many problems on the informational side than they have solved. The results of the third enquiry which covers the whole of rural labour and which was undertaken in two separate parts between 1963 and 1965 are currently in the process of tabulation. It is possible that in the next six months, the tabulation process will be over and the statistical part would be available for interpretation. This, then, will be the main source of carefully checked data for the Commission's report on conditions of rural labour. Apart from the data on employment, the investigations provide interesting insights into the rural consumption patterns and their way of work.

The next source of information to be organised will be the data on consumption patterns in rural areas. Valuable work was done by the Indian Statistical Institute, Calcutta, for the Mahalanobis Committee** which reported in February, 1964 on the distribution of national income in its difficult quest to find out how the additional income generated as a result of 10 years, planning was distributed. It has come to the conclusion, on the basis of data collected through the National Sample Survey (N.S.S.) over a long period of years that in the lowest income bracket the conditions had not improved and it is this bracket in which agricultural labour was one of the main components. The raw material which formed the basis of the conclusions reached by the Income Distribution Committee, will have to be studied in conjunction with the data available since the publication of the Committee's report. The usual sorting from the overall data for rural areas of the statistics which will provide insights into the lower income strata will have to be undertaken. This, then, becomes the second important source on which to base the conclusions on rural labour.

A cross-check of these data will be available through the labours of universities and other research institutions which undertook studies on farm-management. These studies, though they do not have a direct bearing on rural

---

** Report of the Committee on distribution of income and levels of living : Part I—Distribution of income and wealth and concentration of economic power.

labour, provide enough background for understanding how work is organised on farms of different sizes and the economics of such work. Reading these studies particularly with reference to the manner in which small farms are managed, the amount of idle time small farmers have to spend and so on one could build up a comparison with the conclusions of the first and second Agricultural Labour Enquiries. Since these studies have been undertaken in different years, there will be difficulties of interpreting these data consistently. But it is always possible to pick out some threads which may lead to a better understanding of the rural labour situation. It has to be borne in mind that this will be a secondary source of information and due caution will have to be exercised in tying it up with the main source. Of particular interest will be the pattern of work which will indicate with some approximation the state of under-employment of rural labour under different combinations of other inputs.

The Land Reforms Division of the Planning Commission has set up a Central Advisory Committee on Agricultural Labour for drawing up programmes for agricultural labour and for advising State Governments through the Planning Commission on various measures which should be undertaken to improve their lot. The deliberations which have taken place in this Committee over a period of years will be useful in understanding another facet of the problem *viz.*, the amount of land distributed among landless labour in the last 10 years and the future programme in this field. This Committee has, from time to time, prepared an assessment of how the scheme of distribution of land to landless labour has worked. The close association of experienced voluntary organisations and non-officials who have worked in this field for long gives a different flavour to their deliberations. It should be possible to get an evaluation of the various schemes for understanding the extent of the transfer of surplus land from owners to those who had no land. Also useful will be the difficulties experienced in the work of transfers as also the proper upkeep of distributed lands. There is a periodic assessment of the programme of Rural Works which are primarily intended for giving work to labour in slack season, and this could be an important programme to provide adequate work for the landless and building up community associations which help develop agriculture. On the employment side, such assessment has been done by the Directorate General of Employment and Training, and on aspects other than employment, reports will be available from the Programme Evaluation Organisation of the Planning Commission.

The need for a detailed assessment of the changes which have taken place in the conditions of agricultural labour, on the basis of the first two Enquiries, led to a considerable interest in the academic world on problems of comparison of data from two diverse sources. Various ancillary data were pressed into service to understand the real nature of the change. On examining all this material the technical committee, that was set up by the Planning Commission, produced a report which showed areas where matters had improved in the case of agricultural labour and others where they had not. These again raised a public debate till the matter was settled for the time being in a Seminar organised by the Institute of Economic Growth in 1961. The documentation organised for the Seminar will be an important source for a further insight.

The conclusions reached in the Seminar appear to be somewhat obvious, namely, that a solution to the rural problem lies as much in the development of agriculture as it does in the development of sectors other than agriculture particularly, industrial sector. The same theme was developed in a much broader context and academic sophistication in "A Round Table on Economic Growth and Social Justice" organized by the Institute in February 1965. In a learned discussion on the subject a session was specially reserved for problems of rural development. As expected, a prominent theme which was closely debated was the adequacy or otherwise of work on farms and the need for diversifying employment in rural areas by including non-agricultural occupations. Adaptability of rural population particularly engaged in agriculture for work outside their traditional occupations and the need for training was brought out clearly as also the difficulties in the process.

The two important seminars held in recent years are expected to provide valuable documentation on (*i*) the social aspects of the problem of agricultural labour (the Seminar on the Employment of Scheduled Castes and Scheduled Tribes, organised by the Planning Commission in 1964) and (*ii*) an assessment of the policies and programmes for rural labour in the first three plans and the arrangements for the fourth—the first National Seminar on Agricultural Labour in August 1965 organized by the Labour Ministry. These contain information which has considerable analytical value for understanding the problems of agricultural labour. Mainly as a result of the recommendations which emerged out of the discussions on the comparability of the results of the first and second Agricultural Labour Enquiries, case studies of selected villages have been organised by the Labour Bureau Simla along with the Third Rural Labour Enquiry. In terms of assessment of an important piece of labour legislation which affects agricultural workers the recent report by the one man Committee on Minimum Wages throws up problems of fixation of minimum wages in agriculture.

In addition to this, the Commission has on its own initiated during this year a number of, what it calls, "diagnostic studies of problems of rural labour". The plan of these studies is simple though it may not appeal to connoisseurs of statistical methods. A simple form has been drawn up for collection of information in which the Commission is interested ; information is sought to be collected on these forms by properly oriented Education Officers of the Central Board of Workers' Education through discussions, among others, with the village Panchayat officials, leading and progressive farmers, small farmers, landless agricultural workers and Block Development officials. A detailed report based on these studies would be prepared and comments of the State Governments on the same would be obtained. This report with such modifications considered necessary by the State Governments would form yet another basis for deliberations of the Commission. Each State was requested to select some representative districts and in each district villages were chosen taking into consideration (*i*) its nearness to any urban industrial centre and (*ii*) its coverage under developmental activities by rural works, community development and I.A.D.P. Programmes. Normally, five villages are to be selected in each district but it has not been possible to canvass five villages in all the districts. While in some districts five villages have been studied, in others two or three villages alone have been studied. One other important aspect which has been noticed and which is also fairly well-known is the impact of irrigation facilities on the rural employment. Though this aspect has not been specifically taken into account in the selection of the villages, this has been taken into consideration in the analysis of the effects of irrigation facilities in providing better employment opportunities etc. The importance of the classification on the basis of irrigation facilities becomes important because it has been established that where proper irrigation is available the rural employment problem acquires a different meaning. Cropping patterns have undergone a change in favour of crops which require more labour and bullock power or even mechanical power. The result has been that it is the person who can offer himself for work who is in demand and not work for him. Some of these changes have already been written on particularly from the labour angle. These diagnostic surveys which have been completed in almost all the States support the impressions of such transformation and some significant hypothesis can be built thereon.

Apart from these important pieces of evidence there will be replies to the questionnaire issued by the Commission, one section of which is devoted to rural and unorganised labour. While the employers and workers organisations have evinced little interest in this part of the questionnaire beyond saying that something should be done for this class of workers, there is valuable information in the memoranda received from Central and State Governments. This, when analysed, may add to the volume of information discussed above.

It is recognised that all these pieces in the documentation will have their own limitations. There will be obvious dangers of interpreting such disparate data. At the same time the Commission can hope to convey a picture of rural labour and changes which have taken place in it by a careful assessment of facts. Some important threads, it is expected, will still be missing. But in suitable combinations these pieces may present a picture which may approximate to what obtains in reality. If the arrangement of these pieces is again subjected to expert handling in a suitable discussion forum, the Commission may be able to find something by way of a broad consensus of what has happened so far. The task of the Commission, however, does not end there. It has to project into the future on the basis of this experience. For this purpose it is necessary to know about future plans for agricultural development. The Department of Agriculture is preparing a paper on perspective for agriculture and in this perspective it has been requested to indicate the place which agricultural labour will have. Taking advantage of an assurance given in the first National Seminar on Agricultural Labour that the second should be held to assess the effect of recommendations made in the first, the Commission hopes to organise soon this second Seminar to help it in coming to right conclusion.

EXPLORATION INTO THE PRODUCTION ECONOMICS OF HIGH YIELDING VARIETIES

by

W. DAVID HOPPER*

It is not difficult to document the fact that the yields per hectare of India's major foodgrain crops are very low by standards of contemporary agricultural technology. What is of greater significance, however, is the fact that during the past two decades yields have not risen with sufficient rapidity to compensate for the decelerating rate of growth in the land area being put to agriculture and specially to foodgrains, the result has been a steady and observable decline in the rate of growth of agricultural output.

An analysis of production, area and productivity by the three plan periods points dramatically to the role a relatively stagnant expansion of productivity has played in the "agricultural crisis" besetting the country. Table 1 gives the growth rates in per cent per annum (compound) for the three five-year plan periods. The Third Plan period covers only four years, three years when foodgrains production was virtually stagnant at about 80 million metric tons rising in the fourth year to 89 million metric tons. The final year of the plan period, 1965-66, was disastrous for agriculture due to a widespread unprecedented drought caused by a failure of the monsoon rains.[1]

TABLE 1

Growth rate in per cent (Compound) per annum of production, area, and productivity of Indian Agriculture during the periods of the Five-Year Plans, 1951-52 to 1964-65.

Plan Period	*Crop*	*Production (%)*	*Area (%)*	*Productivity (%)*
FIRST PLAN	Foodgrains	6·2	3 3	2·8
(1951-52 to	Non-foodgrains	3·2	2·4	0·7
1955-56)	All crops	5·1	3·1	1·9
SECOND PLAN	Foodgrains	4·2	1·3	2 8
(1956-57 to	Non-foodgrains	3·4	1·3	2·2
1960-61)	All crops	3·9	1·3	2·6
THIRD PLAN	Foodgrains	2·3	0·03	2·2
(1961-62 to	Non-foodgrains	4 2	0·5	3·7
1964-65)	All crops	2·9	0·1	2·8

Note. These rates have been determined by a least squares fit of the relevant indices to the equation form: $I_t = cb^t$ for the years involved. At best, the decimal place is indicative only. Because of the short periods embraced by each rate calculation, each must be regarded as "suggestive".

---

*The Rockefeller Foundation, New Delhi.

[1] If the five years of the Third Plan period are used, the least squares average annual rate of (compound) growth are :

	Production	*Area*	*Productivity*
Foodgrains	—1·5%	—0·9%	—0·6%
Non-foodgrains	1·2	0·2	1·0
All crops	—0·5	—0·6	0·2

The findings presented in Table I suggest that while non-foodgrains productivity has been increasing, the growth in yields of foodgrains has been constant over the 14 years of the three plan periods. Because foodgrains account for two-thirds of aggregate agricultural output, and because productivity expansion has been stagnant while area expansion has slowed, the expansion of total output of grain has slowly declined until it dropped below the 2·4 per cent rate of population growth. This is the crux of the nation's food problem. The focus of concern of this paper, like the central concern of official agricultural policy for the Fourth Five-Year Plan, will be on foodgrain productivity. The narrow purpose of the paper is to explore some of the fundamental agronomic and economic aspects of the new varieties of cereal grains that find pride of place in the Fourth Plan strategy for agricultural development in the "High-Yielding Varieties Program".

There is a mounting body of agronomic evidence to support the view that the biological materials traditionally available to the Indian farmer are closely adapted to his ancient methods of agriculture. Indeed, genetic studies of indigenous varietal materials indicate the generations of perceptive farmers adapting their agriculture to their environment has resulted, over time, in the reduction of the genetic diversity within local varieties to a point where further breeding and selection for higher yields could be productive of only marginal advances.

Into this balanced, adapted agricultural tradition, the quest for higher yields has thrust the inputs of modern agricultural science. The early results of this thrust were not encouraging. Indigenous crop varieties, adapted to the stringent circumstance of Indian agriculture, were found to be unable to respond productively to the new inputs. Varieties that would produce at a low level in the absence of abundant nutrients, produced at approximately the same low level with nutrients or, most serious in wheat and rice, grew rank and tall in a vegetative response only to lodge and lose yield when grain began to form.

Mathematical functions summarizing grain yield response of traditional varieties to nitrogen and phosphate fertilisers for most of India's main foodgrain crops are readily available as a result of the fertilizer trial work of the Institute of Agricultural Research Statistics.[2] These functions were developed from a large number of studies of yield increases of indigenous plant materials when grown with different levels of nutrient application. Almost without exception, the yield of these varieties reached a maximum at nutrient levels below 75 pounds of nitrogen per acre, and below 60 pounds of P_2O_5 per acre. Increases in output even at the maximum levels of nutrient input were severely limited, often less than 1,000 pounds per acre. As a rough aggregate "yardstick" for the response in yield to be expected from

---

[2] See my article on "yardsticks" in the Agricultural situation in India, September, 1965, for a set of fertilizer-yield response functions.

applying nitrogen and/or P_2O_5 to an acre of the local varieties, the official estimate has been ten pounds of additional foodgrains for one pound of nitrogen and six pounds of added grain for one pound of P_2O_5. At the grain prices prevailing in India until about 1963, ten pounds of grain would sell for about Rs. 1·30, six pounds for about Re. 0·78; the farm price for a pound of nitrogen wasabout Re. 0·77 and for P_2O_5, Re. 0·52. The margin between returns and costs for fertilizer was not large, often not large enough to cover normal risks.

The characteristic low yield and the low yield responsiveness of indigenous plant materials to fertilizers, can be traced to several botanical, agronomic, and environmental factors. Chief among these for wheat and rice is the leafy, tall statured, weak-strawed growth habit of the Indian varieties. Jennings[3] has argued that for rice these characteristics can be traced to natural environmental factors conditioning the evolution of the plant. It is his contention, and the contention of other agronomists now working on problems of crop response under tropical conditions, that because of limited supplies of plant food, tropical grain plants have developed an early and profuse leafing habit that allows them to compete for nutrient against weeds and other plants of their own species by mutual shading. This had two adverse effects on grain yield : (1) it reduced plant population per unit area ; and (2) eventually resulted in the plant supporting a lower leaf structure than was too shaded by its own upper leaves to be useful for photosynthesis, indeed, these lower leaves probably respired (by breaking down carbohydrates already synthesized by the plant) more hydrocarbon materials than they added to plant growth by photosynthetic activity and were, therefore, a net drain on the plant at a time when most of the plants reserves of food were needed for grain formation in the reproductive portion of its growth cycle.[4]

Using recently introduced exotic plant materials as a source of increased genetic diversity for plant breeding, and by making direct imports of dwarf wheat and rice varieties from Mexico and Taiwan, Indian Agricultural scientists in the past few years have changed drastically the biological base of the nation's major cereal crops. Varietal banks of worldwide collections for maize, grain sorghum (jowar), and pearl millet (bajra) have been built in India with the assistance of the Rockefeller Foundation. From these materials have come eight hybrid maize varieties, two hybrid sorghums and one hybrid millet. The new hybrids of these three cereals have been augmented by direct imports (and strong local selection and breeding programmes) of dwarf wheat and rice varieties that are non-leafy, stiff-strawed,

---

[3]. Jennings, Peter R. "The Evolution of Plant Type in *Osyza satina L.*" mimeograph, International Rice Research Institute, Los Banos, Philippines, 1965.

[4]. *See : Tanaka, A. Growth Habit of the Rice Plant in the Tropics and Its Effect on Nitrogen Response.* The International Rice Research Institute, Los Banos, Laguna, Philippines, for an excellent discussion of the problem of mutual shading in tropical ricos grown under field conditions,

short-statured plant types that allow multiple tillering and dense plant populations. In all cases the new varieties are considerably more responsive to fertilizer than the local materials traditionally available to the farmers.

It is still too early, however, to say just how responsive to fertilizer these high yielding varieties are. Only maize hybrids have been available for enough time for adequate experimentation. Research has only begun on rice and it is too early to make detailed use of Indian data. The study of wheat is at the stage where research workers are still learning about the proper agronomic practices that must be followed if the full yield potentials of the new plant materials are to be realised. For example, during the rabi season of 1964-65 poor experimental results from dwarf wheats were traced to inadequate control of depth of planting and poor irrigation practices. Experiments during the rabi season of 1965-66 demonstrated that planting depths for these dwarfs should not exceed two inches in contrast to the usual practice with indigenous varieties of placing seed at five to six inches below the soil surface. As a result of the shallower sowing, moisture availability in the top layer of the soil became critical to plant development, and it was found that for best yields, the first irrigation must be given earlier than the customary 40 days after sowing An irrigation experiment recently conducted at the Indian Agricultural Research Institute, New Delhi, indicated that a full irrigation given 20 days after sowing added very nearly one metric ton to yield per hectare, another irrigation at the usual 40 days interval added another metric ton to output ; and because the dwarf wheats are resistant to lodging—they are stiff strawed and too close to the ground for the heavy winds of March to wreck serious damage—they can withstand a late irrigation (say, 130 days after sowing) when the extra water facilitates grain formation and increases yield by an additional 300 to 400 kilograms. Unfortunately, experiment stations are just beginning to conduct complex trials under closely controlled conditions of planting, irrigation and fertilization, and it is likely to be another two to three years before solid data on fertilizer-wheat response is available. Our knowledge of the demands of the hybrid varieties of jowar and bajra is likewise limited, and while the agronomic practices may be less complex, the nuances of their production demands have not yet been explored fully.

While scientists still have much to learn about the new varietal materials, many cultivators are doing as well or better than the professionals. There seem to be several reasons for this. Among them being the fact that many farmers were obtaining higher yields than the experiment stations with the indigenous materials by using husbandry practices that can often be carried over with a few appropriate changes to the production of the exotics. It is, therefore, not too surprising to find that farmer yields on national demonstration trials average almost 1.5 tons per hectare more than experiment stations yields. In 81 national wheat demonstrations laid out in the rabi season of 1965-66, local varieties, that is, indigenous wheats grown under irrigated conditions, averaged 3046 kg/ha as against 4183 kg/ha for the dwarfs, a 37 per cent gain for the imported material. The ranges on these yields over-lapped, a yield range of 1,000 to 4,893 kg/ha for the local varieties and 2,028 to 6,800 kg/ha for the dwarfs. In analysing the frequency distribution of these demonstration results, over 15 per cent of the local yields were below 2,000 kg/ha, none of the dwarf varieties were this low. At the top of the distribution, one-third of the trials with imported materials were over 5,000 kg/ha while only two per cent of the trials with local grain types reached this level. In the major wheat growing areas of the country (the Punjab, northern Rajasthan, western Uttar Pradesh, and Delhi State) the dwarf wheats averaged 5,340 kg/ha against the locals at 3,330 kg/ha, a 60 per cent difference. Experiment station results, on the other hand, seldom exceed 4,000 kg/ha. In explaining this phenomenon, senior research scientists contend that it is easier to take a large farm field and manage it for a high yield than it is to manage a small plot within the constraints of an experimental design. This is particularly true when the optimal management practices needed for one variety are quite different from those needed for another variety, say, irrigation on the dwarf wheats in an experiment that includes a comparison with the tall local varieties placed side-by-side in a single experimental design.

Recognizing the limitations of experimental results, the absence of adequate data for rice under Indian conditions (substituted below are data from the Philippines for Taichung (Native) 1, the main dwarf rice import included in India's High Yielding Varieties Programme) and the fact that very little has been done to ascertain response to nutrients other than nitrogen, the data presented in this paragraph must be regarded only as an indicator of comparative response potential between the exotic and older varieties. The data for local plant materials were gathered from a cross-section of conditions ; as such, the yields quoted as much below those attained by good cultivators enjoying a high level of resources endowment and personal skills. But while absolute yields will vary among farmers, and between the research centres, there is no evidence that response to fertilizer will vary in the same way. Indeed, there is considerable evidence of the opposite. Experimental centres usually have higher initial fertility levels due to a generally more lavish use of plant nutrients, therefore, experimental response trials often show a flatter relation (less yield response to successively higher levels of applied nutrient) than is encountered on farmers' fields. In short, the absolute yield levels outlined below must be considered with caution as there is likely an upward bias for the new varieties and an under-estimate of yield for the local materials, but the shape of the response function itself is likely to be biased in the opposite manner.

For all equations Y is yield of grain, and N is elemental nitrogen applied, both in units of kilograms per hectare.

(*a*) **Rice**

High yielding variety :

Taiwan dwarf indica Taichung (Native) 1

(Philippine data for wet season response)

$$Y = 5330 + 30{\cdot}667N - 0{\cdot}1355N^2$$

maximum yield of 7065 kg per hectare at 113 kg of nitrogen per hectare.

Local varieties (all-India average) :

$$Y = 1840 + 13{\cdot}42N - 0{\cdot}102N^2$$

maximum yield of 2281 kg per hectare at 66 kg of nitrogen per hectare.

Response and yardstick[5]

Nitrogen levels (kg of N/ha)	*New Variety*		*Local Varieties*	
	Total increase (kg of grain/ha	*Yardstick (kg of grain/kg) of N)*	*Total increase (kg of grain/ha)*	*Yardstick (kg of grain/kg of N)*
0 to 20	560	28·0	228	11·4
0 to 40	1010	25·2	374	9·4
0 to 60	1352	22·5	438	7·3
0 to 80	1586	19·8	=	—
0 to 100	1712	17·1	—	—

(*b*) **Wheat**[6]

High yielding varieties – Mexican Dwarf :

(all-India average response from regional trials with Lerma Rojo, aggregated by weighting each regional trial by the irrigated acreage devoted to wheat in the region)

$$Y = 2161{\cdot}7 + 19{\cdot}75N - 0{\cdot}069N^2$$

maximum yield of 3575 kg per hectare at 143 kg of nitrogen per hectare.

Local varieties (all-India average for irrigated wheat)

$$Y = 1455{\cdot}0 + 17{\cdot}20N - 0{\cdot}114N^2$$

maximum yield of 2104 kg per hectare at 75 kg of nitrogen per hectare.

---

[5] Technically the "yardstick" is the area under the marginal physical product curve between zero and the specified upper level of input divided by the upper level of the input. This will give the average increment to total product per unit of input. In mathematical terms it would be $\int_0^b \frac{f'(N)\,dN}{b}$ where $f'(N)$ is the first derivative of the total response function.

[6] The response function for Lerma Rojo was taken from a mimeograph paper by Saxena, P.N. and Sirohi, A.S. "*Responses to Nitrogen with Mexican and Indian Wheats*" Indian Agricultural Research Institute, 1966. The function for the local varieties was taken from the fertilizer response trials conducted by the Indian Agricultural Research Institute.

Response and yardsticks :

Nitrogen levels (kg of N/ha)	New Variety		Local Varieties	
	Total increase (kg of grain/ha)	Yardstick (kg of grain/kg of N)	Total increase (kg of grain/ha)	Yardstick (kg of grain/kg of N)
0 to 20	367	18·4	298	14·9
0 to 40	680	17·0	506	12·6
0 to 60	937	15·6	622	10·4
0 to 80	1138	14·2	—	—
0 to 100	1285	12·9	—	—

(*c*) Maize[7]

Hybrid :

$$Y = 2418 + 22{\cdot}61N - 0{\cdot}04N^2$$

maximum yield of 4974 kg per hectare at 283 kg of nitrogen per hectare.

Local :

$$Y = 2052 + 17{\cdot}23N - 0{\cdot}05N^2$$

maximum yield of 3615 kg per hectare at 172 kg of nitrogen per hectare.

Response and yardsticks :

Nitrogen levels (kg of N/hal)	Hybrid		Local	
	Total increase (kg of grain/ha)	Yardstick (kg of grain/kg of N)	Total increase (kg of grain/ha)	Yardstick (kg of grain/kg of N)
0 to 50	1031	20·6	737	14·7
0 to 100	1861	18·6	1223	12·2
0 to 125	2200	17·6	1373	11·0
0 to 150	2491	16·6	1460	9·7

(*d*) Jowar[8]

(The data for sorghum, both hybrid and local are very crude and incomplete. What is presented below should be regarded as only most tentative)

---

[7] Data for both the hybrid and local response is from Shah, V.H. "*An Analysis of Response of Hybrid Maize to Nitrogen*", Mimeo Indian Agricultural Research Institute, 1966.

[8] Data for the hybrid response are from the sorghum trials conducted as part of the Indian Council of Agricultural Research programme for sorghum improvement. Data for local varieties are the fertilizer response trials conducted on farmers' fields by the Institute of Agricultural Research Statistics.

Hybrid :

$$Y = 1635 + 22{\cdot}33N - 0{\cdot}0777N^2$$

maximum yield of 3239 kg per hectare at 144 kg of nitrogen per hectare.

Local :

$$Y = 1145 + 9{\cdot}39N - 0{\cdot}0574N^2$$

maximum yield of 1529 kg per hectare at 82 kg of nitrogen per hectare.

Response and yardsticks

Nitrogen level (kg of N/ha)	*Hybrid*		*Local*	
	Total increase (kg of grain/ha)	*Yardstick (kg of grain/kg of N)*	*Total increase (kg of grain/ha)*	*Yardstick (kg of grain/kg of N)*
0 to 40	769	19·2	284	14·2
0 to 80	1289	16·1	384	4·8
0 to 120	1561	13·0	—	—

(*e*) **Bajra**[9]

(The caution set forth in the section on jowar applies even more forcefully to data for pearl millets.)

Hybrid

$$Y = 1496 + 15{\cdot}13N - 0{\cdot}0281N^2$$

maximum yield of 3533 kg per hectare at 269 kg of nitrogen per hectare.

Local

$$Y = 1118 + 9{\cdot}10N - 0{\cdot}02N^2$$

maximum yield of 2153 kg per hectare at 227 kg of nitrogen per hectare.

Response and yardsticks

Nitrogen levels (kg of N/ha)	*Hybrid*		*Local*	
	Total increase (kg of grain/ha)	*Yardstick (kg of grain/kg of N)*	*Total increase (kg of grain/ha)*	*Yardstick (kg of grain/kg of N)*
0 to 50	686	13·7	405	8·1
0 to 100	1232	12·3	710	7·1
0 to 150	1637	10·9	915	6·1

---

[9] Data for bajra, local and hybrid, was taken from Mutry, B. R. "*Fertilzer Response to Hybrids of Jowar and Bajra in 1965*", Mimeo., Indian Agricultural Research Institute, 1966.

Of the five cereals for which there are now new responsive varietal materials available, the hybrids of maize, jowar, and bajra give substantially larger yields than local varieties when no nitrogen is applied. Jowar hybrids averaged 1635 kg per hectare at zero nitrogen, against 1145 kg for the local materials, a difference of over 42 per cent. Unfertilized bajra hybrids also out yielded the domestic types by 1496 kg per hectare to 1118 kg., a difference of 34 per cent. Maize showed an almost 400 kilo-gram difference in yield between the hybrid and local materials at a zero nitrogen application. In the case of wheat and rice, the data are too dis-similar to draw even tentative conclusions, but a few incomplete experiments suggest that on irrigated land, at least, the exotic materials seem capable of giving considerably better yields than the indigenous varieties at low levels of fertility.

In examining the economic returns to the two types of plant material, it is of interest to note that against average incremental returns (yardsticks) of over 15 kilograms of grain per kilogram of *N* for the exotics and 8 to 10 kilograms of grain per kilogram of *N* for the locals, the present farm price of nitrogen is Rs. 2·15 per kilogram, the approximate equivalent of three to four kilograms of grain. At these price relations, and with the new highly responsive varieties, the demand of farmers for exotic seeds and fertilizer has received an accelerative impetus that completely negates the validity of any trend estimate of national fertilizer consumption. At present prices, and allowing for a substantial risk discount, the optimal dosage of nitrogen for the high yielding dwarf wheat is around 120 kilograms per hectare, for wet season rice it is probably about 70 kilograms per hectare and for rice grown during the dry season (when there is a higher sunlight intensity and, therefore, greater photo-synthetic activity and a greater response to plant nutrients) it is about 100 kg per hectare. For the more vegetative crops such as maize, sorghum and millet, dosages of 100 to 130 kg of nitrogen per hectare are probably close to optimum. Research on phosphorous, alone and in combination with nitrogen, is still tentative but what evidence there is indicates a substantial response to this nutrient as well. There are few data on potash response, but what are available suggest that in most of India there is little need for this plantfood.

At an input of 120 kg of nitrogen per hectare to the dwarf wheats, yield would be increased by about 1376 kg. At present farm prices of around Rs. 75·00 per quintal (100 kgs), this extra yield would sell for about Rs. 1032·00, a gross return gained at a cost of Rs. 258·00 for the nitrogen. Even on local wheats a "recommended" dose of 40 kg of nitrogen per hectare would add 506 kg of grain to yield and give a monetary return of Rs. 380·00 for an outlay of Rs. 86.00. (It is of interest to note that at prices prevailing until mid-1963, the local varieties would provide a gross of Rs. 200.00 at a cost of Rs. 70·00, not a highly attractive rate of return when discounted for risk.) It seems clear that under present price relations even the limited response to nitrogen of the older plant types provides an attractive opportunity to invest in fertilizer. It should be anticipated that today's demand for fertilizer for wheat will come from all cultivators having access to irrigation, not merely from those who can procure seed of the high yielding varieties. While less easily documented, a similar statement would hold for farmers growing any of the other crops under adequate moisture conditions.

Under the High Yielding Varieties Programme, the Ministry of Food and Agriculture hopes that by 1970-71, 32·5 million acres will be producing crops from the hybrid and exotic cereal plant materials. Table II gives an approximate estimate of past national output and acreages of production for each of the five crops being promoted in this programme.

TABLE II

Output and acreage for five major cereals, and the acreage to be included under the high yielding varieties programme

Crop	*National output of selected cereals and per cent of total foodgrains production*		*Acreage and per cent of total acreage devoted to selected cereals*		*Acreage called for under HVP by 1970-1971*	*HVP acreage as per cent of normal acreage in each crop*
	(Metric ton)	%	(metric acres)	%	(metric acres)	%
Rice	38·0	42	85·0	30	12·5	15
Wheat	13 0	14	33·0	12	8·0	24
Jowar	10·0	11	45·0	15	4·0	9
Maize	4·5	5	11·0	4	4·0	36
Bajra	4·0	4	27·0	9	4·0	15
Total all foodgrains	90·0		295·0		32·5	11

The fine selected cereals account for about 90 per cent of national cereal production, and about three-quarters of the annual domestic supply of foodgrains.

When the high- yielding varieties programme was first formulated, it was thought that the major constraint on area growth would be seed availability. It appears now that seed will not be a constraint on expanding the acreage under wheat and rice. The area under these crops will be limited mainly to land that is adequately watered and shaped to allow some element water management. For the three hybrids, the availability of quality seed could act as a serious constraint on the extension of acreage.

The most significant constraint on realising the yield potential of the new varieties (in contrast to area expansion) is likely to be a shortage of fertilizer. At application rates for nitrogen given above, the acreage target for the high yielding varieties programme would require 1·3 million tons of nitrogen, 65 per cent of the 2·0 million tons projected as being available in 1970-71. In other words, if the programme succeeds in meeting its planned acreage, and if these acres are fertilized at what appear to be optimum rates from the point of view of the farmer[10], 65 per cent of the nation's available nitrogen will go to about 11 per cent of the national acreage devoted to foodgrains. No projections are easily possible for phosphate requirements. But if P_2O_5 applications were to average 50 kg per hectare (the presently recommended rate for many areas of the country), requirement would be 658,000 tons of P_2O_5 or, again, about 65 per cent of the 1·0 million tons of planned availability.

It is important to note that the higher yield obtained on the new varietal material is not merely due to the fertilizer uptake of the plant. Indeed, it is due mainly to the genetic ability of the variety to yield under conditions of crowding, *i.e.*, to produce under conditions of a high plant population per hectare. High plant populations require high fertility soil, and as the larger leaf area per hectare of the increased number of plants involves a larger transpiration surface, there is a substantially greater use of and, need for water than is the case with traditional plant types. It has already been mentioned that for high yields from the dwarf wheats it is necessary to give close attention to irrigation and irrigation timing (a matter that has far-reaching implications for the scheduling of irrigation water in the canal areas, and in part, accounts for the very strong demand for private irrigation systems of modern design that free the owner from a dependence on governmental sources that are operated to meet the agronomic demands of the older plant materials) but for all crops the water up-take of the new varieties will be substantially greater than for the older types, a factor that will impose a limitation on their general adaptability by holding them to areas with either irrigation or reasonably assured rainfall during the growing season. Estimates of how many acres there are in India presently adapted to the availed responsive varieties are very crude, but it seems likely that 12 million acres could be planted successfully to wheat, about 35 to 40 million acres to the three hybrids, and between 20 to 30 million acres to rice. These estimates do not include double cropping potential. Already, in parts of the north-Indian plains, farmers are experimenting with hybrid maize-wheat, and hybrid jowar-wheat rotations. The timing of these rotation sequences

---

[10] The application of nitrogen per hectare needed to maximize cultivator profit is not necessarily optimal for gaining maximum total additional production for the nation as a whole when there is a limited fixed quantity of nutrient available for the country. But it is hard to see how farmers can be effectively prevented from selecting and using personally profitable application rates for fertilizer even if these entail a social cost and, likely a gray market in the scarce input. In fact, it is likely that the social outlay needed to prevent the pursuit of private gain would be more than the added return from a more efficient (social) use of fertilizer, and such preventive measures could very well remove any farmer incentive to innovate by discouraging a change to the new varieties and new methods. For the social argument see : Minhas, B.S. and Srinivasan, T.N. "New Agricultural Strategy Analysed" *Yojana* January 26, 1966.

involves a critical period of land preparation for wheat sowing that is not handled easily by bullock tillage power. As a result, demand for farm tractors has increased and will likely become a major input need pressing close behind the need for irrigation pumps, water spreading equipment, and fertilizer.

The high-yielding varieties have still other consequences that will influence the future technical requirements of Indian agriculture. Larger plant populations lead directly to greater insect populations, and provide an ideal environment for the spread of diseases. With traditional materials, systematic plant protection measures of a prophylactic nature were only marginally profitable, if profitable at all. Relative to cost there seemed little to be gained by adding ten per cent to 1,500 pound per acre yield. But a similar ten per cent on a 4,000 pound crop can pay for a comprehensive plant protection programme and leave a substantial and attractive residual. A similar argument would lead to the conclusion that land shaping for improved water control and water management will become more profitable as percentage losses to low yield crops due to uneven field surfaces are applied to crops of much higher yield capacity. In fact, there is already a substantial demand from farmers for more power, land planes, and even custom services for larger scale earth moving to take advantage of the economic gains from improved opportunities for better water management.

The experimental study of the new varieties is still too sketchy to do more than provide a vague outline of their potential, but this outline suggests strongly that there is a wide gulf between the new and old genetic base on which India's crops can rest. A development strategy for agriculture that is based on the new genetic potential offers a substantial promise of revolutionizing the nation's agriculture. The revolution will not be completed easily, however, for like all beginnings to revolution, the implications and ramifications are more far reaching than those immediately apparent at the start. The new varieties bring with them a set of unadvertised demands for technical changes in Indian agriculture that will have far-reaching consequences not only on the rural economy, but on the industrial base of the country, on the national infra-structure, and on many of the traditional shibbolths of Indian governmental administration. The use of these responsive plant materials, and the support of their use by an economy and an administration geared to promote agricultural development, will ultimately pose significant problems of social equity. The new varieties are best suited to areas already advantaged with irrigation and assured rainfall. Their yield potential is of an order that means the benefits they confer on the farmers in these advantaged areas cannot be ignored in a development process that holds the social equalization of income as one of its goals. Indeed, the implications and opportunities the new materials have for the political economy of the whole country are undoubtedly to be their most important continuing legacy. But this is another story.

A GENETIC ANALYSIS OF YIELD AND SOME FIBRE PROPERTIES IN GOSSYPIUM ARBOREUM L.

by

A. B. JOSHI, MUNSHI SINGH, S.N. KADAPA*

Introduction. Among the species of cultivated cottons grown in India, the *desi* cottons belonging to the diploid species, *Gossypium arboreum* L. and *G. herbaceum* L., occupy 61 percent of the total acreage and contribute to 64 per cent of the total lint production. The remaining area under cotton is planted with the tetroploid American cottons (*G. hirsutum* L) ; the area under *G. barbadense* L is almost negligible (4200 hectares in 1964-65). The total acreage under all the species of cotton in India is about 8·1 million hectares and the lint production about 5·4 million bales of 180 kg. each. Although the majority of the *desi* cottons comprise short to medium staple types, they are hardy, disease and pest resistant and are extensively grown under rainfed conditions. On the other hand the tetroploid cottons, which are primarily grown for their superior fibre quality, can be successfully cultivated only under irrigated or assured rainfall conditions and have to be protected against pests and diseases. The improvement of *desi* cottons with regard to yielding ability and fibre quality is, therefore, important from the stand point of cotton economy in the country.

The length, fineness and tensile strength of the fibre constitute the major characters influencing the spinning performance of cottons. Therefore, in the genetic improvement of this crop it is important to combine superior fibre quality with increased yielding ability. All these characters are quantitatively inherited and, therefore, the cotton breeding procedures must be based on a sound knowledge and understanding of their mode of inheritance and their inter-relationships. Panse along with J.B. Hutchinson and his other colleagues at Indore were among the pioneers in the field of cotton breeding and genetics in India.

While a considerable amount of literature has accumulated on the inheritance of fibre quality in *G. hirsutum* L and *G. barbadense* L., researches in this direction in *G. arboreum* L. and *G. herbaceum* L. have been rather meagre. The studies of Hutchinson, Panse and Govande (1938), Koshal, Gulati and Ahmed (1940), Govande and Joshi (1950), Kalyanaraman, Santhanum and Ramachandran (1956) suggested the Mendelian type of inheritance for the different characters contributing to fibre quality in *G. arboreum* L. So far, no attempt seems to have been made in this species to subject the characters to quantitative genetic analysis following the recently

---

*Indian Agricultural Research Institute, New Delhi.

developed procedures in statistical genetics. The present paper records such studies on the P, F_1, F_2, B_1 and B_2 generations of an inter-varietal cross, Sanguineum × Virnar, of *G. arboreum* L., with regard to yield, halo-length, fibre fineness and strength.

Material and Methods. Sanguineum and Virnar were used as the parental varieties for this investigation conducted during 1965-66. Sanguineum is a high yielding, short stapled land variety of *arboreum* cotton, with a wide range of adaptation in the northern cotton growing States of India. Virnar is an improved strain developed in the erstwhile Bombay State and released for general distribution in 1949-50 (Sethi, 1963). According to the classification adopted by the I.C.C.C. in 1964 it is classed as a medium staple cotton.

The parental, F_1, F_2 and B_1 and B_2 generations of the cross, Sanguineum × Virnar, were grown in 1965, in a randomised block design replicated four times at the Indian Agricultural Research Institute, New Delhi. The plots under parent and F_1 comprised a single row each per replication while the F_2s and the backcross generation material had each four rows per replication. The rows were spaced 2 ft. apart with 1 ft. distances between plants in the rows. In each case there were 15 plants per row.

Observations were recorded on all the experimental plants, with regard to yield per plant, halo-length (mms.), fibre fineness (Micronaire value) and fibre tensile strength (Pressley strength index at zero gauge). The analysis of variance was worked out, as usual, on the basis of plot means. Intrarow variability, taken over all the four replications, was calculated for each generation and the standard deviations of generation means were also worked out.

The total variance was split into genotypic, phenotypic and environmental components for each character, worked out by the method given by Mather, 1949. The six genetic parameters *viz.* m, d, h, i, j and l, were estimated on the lines suggested by Hayman (1958) with the help of generation means and the standard deviation of each component was calculated on the basis of inter-plant variability in each generation.

Results. The mean values for yield and the three fibre characters are presented in Table 1. The two parents used in the cross exhibited wide differences in all the four characters studied. Sanguineum was significantly superior in yield per plant to Virnar, while the latter was superior in respect of the three fibre characters. The differences were significant statistically at the 1 per cent level. The F_1 value showed that high yield of *kapas* (seed-cotton) per plant was completely dominant over low yield while the mean halo-length did not deviate significantly from the mid-parent value ; higher micronaire value (coarseness) and lower fibre strength were found to be

TABLE 1

Mean values for yield and fibre characters in the parental F_1, F_2 and back-cross generations.

Generation	**Kapas** *yield per plant (gms)*	*Halo-length (mms)*	*Fibre fineness (Micronaire)*	*Tensile strength. (Pressley index)*
P_1—Sanguineum	37·27	14·40	7·30	5·50
P_2—Virnar	13·75	21·60	4·60	7·25
Mid-parent value	25·51	18·00	5·95	6·37
F_1	37·75	19·32	6·44	5·92
F_2	30·45	18·12	6·32	5.76
B_1—(F_1×Sanguineum)	39·05	16·58	6·75	5.63
B_2—(F_1×Virnar)	25·60	20·71	5·72	6·14
S.E.M.	1·471	0·380	0·129	0·226
C.D. at 5%	3·678	0·950	0·422	0·566
C.D. at 1%	4·895	1·129	0·562	0·753

partially dominant. The depression from the F_1 to the F_2 generation in respect of yield of *kapas* per plant was of the order of 19·28 per cent, but the F_2 mean was higher than the mid-parent value. The F_2 mean did not significantly deviate from the mid-parent value with regard to halo-length but it was higher than the respective parent with regard to coarseness and lower strength of fibre. The mean yield of B_1 (*i.e.* back-cross of F_1 to Sanguineum) was higher than the Sanguineum parent, although not significantly. In respect of halo-length, fibre-fineness and strength the mean of backcross progenies was at the expected level between the F_1 and the respective parents.

The estimates of the six genetic parameters, as per the notation of Hayman (1958), are given, along with their standard deviations, in Table 2. For yield, the mean genetic effects in F_2 (m) was high (30·45) with a rather large standard deviation suggesting that selection for genetically-superior-yielding progenies should be effective in the F_2. The dominance effects of genes, (h) appear to contribute more (19·74) than the additive genetic effects (13·45). On the other hand, both additive and dominance gene effects appeared to be equally responsible in the control of halo-length and fibre strength. In the expression of fibre fineness, mainly additive effects appeared

TABLE 2

Genetic parameters by "generation mean metbod" of yield and fibre characters

Genetic parameters	*Kapas yiel.l per plant*	*Halo-length*	*Fibre-fineness*	*Fibre-strength*
m	30·45 (±11·93)	18·12 (± 2·12)	6·32 (± 0·70)	5·66 (± 0·04)
d	13·45 (± 2·68)	4·148 (± 2·53)	1·03 (± 0·05)	0·50 (± 0·07)
h	19·74 (± 7·10)	3·40 (± 10·0)	0·15 (± 0·02)	0·44 (± 0·05)
i	3·75 (± 6·90)	2·08 (± 9·90)	−0·34 (± 0·02)	0·898 (± 0·14)
j	3·38 (± 3·79)	1 22 (± 5·29)	0·64 (± 0·13)	0·75 (± 0·17)
l	−10 28 (± 9·25)	−2·00 (± 13·6)	0·08 (± 0·29)	0·15 (± 0·16)

to be operative. The magnitudes of the *i* (additive × additive) and *j* (additive × dominant) types of interaction were approximately of the same order for yield per plant and fibre strength, but the *i* type of the gene interaction was preponderant in respect of halo-length. The *j* type interaction appeared to be more important with regard to fibre fineness. The dominant × dominant type of interaction (*l*) was either to small to be effective in the control of fibre strength and fineness or had a negative effect in the expression of yield per plant and halo-length.

Phenotypic correlations between the four characters were studied in the parental, F_2 and back-cross generations (Table 3). Yield of *kapas* per plant,

TABLE 3

Phenotypic correlations between *kapas* yield and fibre characters in the parental, F_2 and backcross generations.

	Halo-length	*Fibre-fineness*	*Fibre-strength*
Kapas yield P	−0·647**	0·470**	−0·747**
F_2	0·248	0·191	0·293*
B_1	−0·0014	0·023	−0·005
B_2	0·039	0·059	0·254*
Halo-length P	—	−0·947**	0·941**
F_2	—	0·668**	0·656**
B_1	—	0·525**	0·676**
B_2	—	0·251*	0·143
Fibre-finenes P	—	—	−0·967**
F_2	—	—	−0·450*
B_1	—	—	−0·500*
B_2	—	—	−0 912**

Values of r for 5 per cent and 1 per cent level of significance are 0·2500 and 0·3248 respectively (Fisher and Yates, 1943).

* : Significant at 5% level.

** : Significant at 1% level.

in the parents was negatively correlated with halo-length (–0·647) and fibre strength (–0·747) but its relation with fibre fineness was rather low (0·470). In the F_2 generation, however, yield per plant was positively related with the fibre characters, though this association was of a low magnitude. The correlation between yield and other characters was almost negligible in the backcross generations. Halo-length was negatively correlated with fibre fineness both in the parental (–0·947) and F_2 generations (—0·668) but when the F_1 was back-crossed to Sanguineum this relationship changed its sign, being positive (0·525). In the reciprocal backcross (to Virnar), the value of correlation coefficient was low (0·251). The relationship of halo-length with fibre strength in all the generations was positive and high except in the backcross to Virnar generation when the '*r*' value was rather low. The relationship between micronaire value and fibre strength was negative and moderate to high in all the generations. In other words, higher fibre strength was associated with lower micronaire value (fineness).

Discussion. In the data presented in this paper an attempt has been made to conduct a complete genetic analysis of yield and three among the most important fibre characters in *G. arboreum* L.

Yield of *kapas* per plant was found to be governed by dominant gene action as the F_1 mean yield per plant significantly differed from the mid-parent value and secondly the *h* (dominance) component of genetic effects were predominant than the *d* (additive) component. Turner (1953) also came to similar conclusions working with *intra-hirsutum* crosses while Stroman (1961) found over-dominance to be important in determining the F_1 yields in intra-*barbadense* crosses. The redeeming feature, in the present case however, is that additive gene effects also contribute substantially to the expression of this character. Secondly, since the F_2 mean yield is above the mid-parental value, it would appear that a major portion of the F_1 yield is contributed by additive gene action. Therefore, selection for increased yield in the segregating generations should be effective.

Koshal *et al* (1940) observed heterosis for fibre length to a considerable extent in the three intra-*arboreum* crosses studied by them. Balasubramaniyan and Santhanum (1952) studied the inheritance of staple length in crosses involving a short-lint (5 to 10 mm.) and a lint-less mutant of *G. arboreum* L obtained in F_2 a 9 : 4 : 3 ratio in the crosses between the two mutants and a 3 : 1 ratio in cross of the short-lint mutant with the normal linted strain, *N* 14 (19·4 mm.). Stroman (1930, 1961), Ramey (1960), Stith (1956), Kamel (1959) Kamel and Asmail (1966) reported that in the intra-*G. hirsutum* and intra *G. barbadense* crosses studied by them longer staple length was partially dominant. In the present study, however, it appeared that halo-length was primarily under additive genetic control, although dominance effects also contributed to it in a substantial measure. It should be therefore possible to exploit the additive genetic effects in selecting for increased staple length.

Fibre fineness was found to be almost fully under the control of additive genetic effects. Koshal *et al* (1940) and Hancock (1944) came to similar conclusion from their studies on intra-*arboreum* and intra-*hirsutum* crosses, respectively. In the present study additive genetic component (*d*) for this character appeared to be considerably high as compared to the three epistatic components. Thus, for this character also selection should results in the isolation of superior strains.

However, the situtation with regard to the mode of inheritance of fibre strength did not appear to be so straightforward. Both the *i* (additive × additive) and the *j* (additive × dominance) types of gene interactions appeared to benstitute substantially, and almost equally, to inheritance of this character. The additive component (*d*) and the dominant component (*h*), however, appeared to be quite important. In *G. hirsutum*, Stith (1956) observed partial dominance for fibre strength, while Dark (1962) found its inheritance to be complex.

A study of correlations between the characters under consideration revealed the existence of strong positive association between *kapas* yield per plant, on the one hand, and halo-length, fibre fineness and strength, on the other. Al-Jibouri, Miller and Robinson (1958) and Miller and Rawlings (1967) found yield and fibre strength to be negatively related in *G. hirsutum L.* although they could not explain why it should be so. Miller, Williams, Robinson and Comstock (1958) did not notice any correlation between fibre length and fineness and strength, in another material of *G. hirsutum* L.

Constant and moderate to high negative correlations found between fibre fineness and fibre strength in all the generations suggests that, although the inheritance of fibre strength is complicated by inter action effects of genes, selection for fibre fineness may lead to a satisfactory improvement of fibre strength also. Since increased fibre fineness, *i.e.*, lower micronaire value, is negatively associated with increased fibre strength even in the segregating generations, it would be apparent that gene segregation does not greatly disturb the desirable negative association between fibre fineness and strength.

The results presented above would suggest that, in this cross combination, yield of *Kapas* per plant is controlled both by additive and dominance effects of genes, while fibre length and fibre fineness are largely under the control of genes whose effects are additive. The genetic control on the other hand, for fibre strength appeared to be rather complex with epistatic (non-alletic interaction) gene effects playing a major role. In view of the fact that fibre fineness and length are additively inherited and yield, predominantly by dominance effect, selection for high yield may not yield segregans superior in respect of fibre fineness and length and *vice versa*. Therefore norms would have to be fixed for all the three characters for selection to be effective. Since fibre fineness and fibre strength showed significant association in all the generations studied selection for fibre fineness should yield segregants with a satisfactory measure of fibre strength also.

However, a word must be said here about the breeding methodology which must be followed for the achievement of success in such a multi-directional selection programme. Obviously, the traditional pedigree method of selection would not be effective here (Joshi 1963 ; Joshi and Dhawan, 1965) and procedures involving *inter se* crossing among superior segregants must be adopted for accumulating together the additive genetic effects. Such *inter se* crossing between selected lines helps in increasing the frequency of genes and gene complexes of the desirable types whose effects are additive in nature, (1960). Pandya, Mujumdar, and Desai (1958) carried out inter-strain crossing among selected strains of *G. herbaceum* L., and Murthy, Rao, Reddy and Ali (1964) did the same in *G. arboreum* L., for combining together more than two desirable quantitative traits. It is however suggested here that such *inter se* hybridization should be taken up before the characters get fixed in different lines and, preferably, the programme should be based on the genetic information obtained from biparental crosses carried out in the F_2 or F_3 generation.

Hanson (1959), on theoretical considerations, suggested that, under the influence of selected mating between several individuals, initial linkage blocks would be broken up and more and more desirable gene combinations would result through each cycle of inter-crossing between selected lines. This suggestion was put into practice by Hanson, Probst and Caldwel (1967) in soyabean crosses where substantial increase in seed-oil percentage and yield of seeds was obtained through four cycles of inter-crossing. Miller and Rawlings (1967) subjected *G. hirsutum* intra-varietal cross progenies to five cycles of natural crossing (50 per cent out-crossing block) and obtained increase in yield and fibre characters.

Although the *desi* cottons belonging to *G. arboreum* and *G. herbaceum* cottons and normally recognised as short and medium staple types, breeding for longer staple, up to as much as 29 mm. has already been done successfully (Bederkar, 1957 ; Pandya *et al.* 1958). If the researches in this direction are stepped up and carried out even more systematically, using the *inter se* crossing procedure discussed above it should be possible to evolve high yielding strains of *desi* cottons possessing far better fibre properties and spinning quality.

Summary : Parental, F_1, F_2, B_1 and B_2 generations of an inter-varietal cross, Sanguineum × Virnar, of *G. arboreum* L. were studied for yield of *kapas* (seed cotton) per plant, halo-length, fibre fineness and fibre strength. In the inheritance of yield both additive and dominance gene effects were noticeable while halo-length and fibre fineness were predominantly under additive genetic control. In the case of fibre strength the additive genetic component (d) was small in magnitude and the interaction effects (i and j) were preponderant. A breeding procedure *inter se* hybridization between desired segregants has been suggested for securing lines combining a good measure of the quantitatively inherited traits under study.

REFERENCES

Allard, R. W. 1960. Principles of plant breeding. Toppan Co, Ltd. Tokyo, Japan, pp. 485.

Al-Jibouri, H.A., Miller, P.A., Ribson, H.F., and Comstock, R.E. 1958. Genotypic and environmental variances and covariances in an Upland cotton cross of inter-specific origin. Agron. J. **50**, 633-638.

Balasubramaniyan, R., and Santhanum, V. 1952: Inheritance of short lint mutant in Coconada cotton. Cor. Sci. **21**, 16-17.

Bederkar, V.K, 1957. Scope for further improvement of Hyderabad Gaorani cotton. I.C.G.R. **9**, 346-379,

Dark, S.O.S. 1962. Breeding increased lint strength in Sakel type cotton. Emp. Cott. Gr. Rev. **39**, 161-169.

Govande, G.K. and Joshi, N.V. 1950. Inheritance of Agricultural characetrs in three inter-strain crosses in cotton. I.C.G.R. **4**, 46-54.

Hancock, N.L. 1944. Fineness and strength of cotton lint as related to heredity and environment. J. Am. Soc. Agron. **36**, 530-536.

Hanson, W.D. 1959. Break up of initial linkage blocks under selected mating systems, Genetics **44**, 857-868.

Hanson, W.D., Probst, A.H., and Caldwel, B.E. 1967. Evaluation of population of soyabean genotypes with implications for improving self-pollinated crops. Crop. Sci. 7, 99-103.

Hayman, B.I. 1958. Separation of epistatic from additive and dominance variation in generation means. Heredity. **12**, 371-391.

Hutchinson, J.B., Panse, V.G. and Govande, G.K. 1938. Studies in plant breeding technique, IV. The inheritance of agricultural characters in three inter-strain crosses in cotton. Indian J. Agric. Sci. **8**, 757-776.

Joshi, A.B. 1963. Production breeding. Indian J. Genet. **23**, 109-116.

Joshi, A.B. and Dhawan, N.L. 1966. Genetic improvement of yield with special reference to self-fertilizing crops. Indian J. Genet. Special Symposium number. **26**, A. 101-113.

Kalyanaraman, S. M. and Santhanum, V. 1957. A review of recent trends in cotton breeding technique, I.C.G.R. **11**, 131-135.

Kamal, S.A.L. 1959. Inheritance of fibre fineness in inter-specific hybrid. G. hirsutum ×G. barbadense. Diss. Abs. **20**, 3-4. (From Pl. Br. Abs. **30**, 3, 570, 1960)

Kamal, S.A.L. and Asmail. A.A. 1962. Inheritance of fibre length and fineness of Egyptian cotton. Emp. Cott. Gr. Rev. **43**, 207-217.

Kosbal, R.S. Gulati. A.N., and Nazir Ahmed 1940. The inheritance of mean fibre length, fibre weight per unit length and fibre maturity of cotton. Indian J. Agric. Sci. **10**, 975-989.

Mather, K. 1949. Biometrical Genetics. A study of continuous variation. Dover publication Inc. USA pp. 158.

Miller, P.A., Williams, J.C. (Jr.), Robinson, H F., and Comstock, R.E. 1958. Estimates of genotypic and environmental variances and covariances in Upland cotton and their implications in selection. Agron. J. **50**, 126-129.

Miller, P. A. and Rawlings, J.O. 1967. Break up of initial linkage blocks through inter mating in a cotton breeding population. Crop Sci. **7**, 199-204.

Pandya, P.S., Mujumdar, P.N., and Desai, K.B. 1958. Breeding for fibre length in herbaceum cotton. I.C.GR. **12**, 77-79.

Ramey, H.H. 1960. Evidence for gene interactions in the inheritance of lint length in Upland cotton. Genetics **45**, 1007-100.

Satyanarayana Murthy, K., Dhananjaya, Rao, V., Narayan Reddy, N.S. and Mahaboob Ali., S. 1964. Evolution of long staple *desi* cotton type under rain fed conditions in Northern tract of Andhra Pradesh. Andhra Agric. J. **11**, 42-49.

Sethi, B L. 1963. A guide to Indian cottons. I.C.C.C. Bombay, pp. 64.

Stith, L.S. 1956. Heritability and inter-relationships of some quantitative characters in a cross between two varieties of *G. hirsutum* L. Iowa Sta. Col. J. Sci. **30**. 439-440.

Stroman, G.N. 1930. Biometrical relationships of certain characters in Upland cotton. J. Ameri. Soc. Agron. **22**, 327-340.

Stroman, G.N. 1961. An Approach to hybrid cotton as shown by intra and inter-specific crosses. Crop Sci, **1**, 363-366.

Turner, J.H. (Jr.) 1953. A study of heterosis in Upland cotton Agron. **J. 45**, 484-486.

IMPACT OF YIELD-INCREASING TECHNOLOGY ON FARM LABOUR USE IN I.A.D.P. DISTRICT, LUDHIANA

by

A.S. KAHLON & TILAK RAJ KAPUR*

Introduction

Introduction of capital-deepening technology is not suitable to the present stage of agricultural development in India. On the other hand, intensive use of labour without any capital intensification will not take Indian agriculture vary far. The pertinent question therefore is : What form of technology would be most appropriate for an overpopulated country like India, where agricultural labour is relatively abundant.

Some innovations have labour using bias and lead to labour absorption without much increase in the stock of capital. This process is better known as capital-shallowing and has characterized the growth pattern of Japanese economy from 1890 through 1930. As a contrast, no capital-shallowing took place in India during the decade 1949-1960. Rather, the capital-labour ratio increased in the industrial sector.

The Japanese experience and the Indian cross section data also indicate that the Indian agriculture can, and probably will, move in the direction of a more labour intensive production function for some time to come. Thus the special emphasis need be laid on the introduction of such technological practices as are complementary to labour use rather than substituting for human labour. It is only after a definite stage of commercialization of agriculture is reached that there will be a need for capital-deepening technology as a labour saving device, consistent with the requirements of rapid economic growth to suit that particular stage of economic development.

Recent introduction of such technological improvements as high yielding mexican wheats, hybrid bajra, hybrid maize, paddy T.N. 1 and the increased use of associated inputs such as fertilizers and plant protection measures have necessitated the use of additional labour inputs, along with increased demand for working capital. This type of technology is of the nature of capital broadening rather than capital deepening type and is more appropriate to the conditions of Indian agriculture.

The present study is an attempt to examine analytically, the extent of additional labour requirements, when various farm situations are reorganized, using different levels of technology.

---

*Punjab Agricultural University, Ludhiana.

More specifically, the objectives of this study are :

(*i*) Appraisal of the labour use pattern on selected farm situations with their existing techniques of production.

(*ii*) Determination of the extent of improvement in labour use pattern through the re-organization of the farm resources based on :

(*a*) existing production techniques.

(*b*) improved production techniques.

Methodological approach

The study was conducted in 'Upper Dhaia' region of the I.A.D.P. District of Ludhiana. Dehlon block formed the operational area of the study because it represented farming characteristics of 'Upper Dhaia' region. A list of operational holdings was prepared in five villages of the block selected at random. These holdings were pooled and distribution transformed to obtain small, medium and large size holdings. Frequency distribution of each size group gave 10 acres, 16 acres and 22 acres as size of holdings for modal class groups, representing small, medium and large groups respectively.

Three representative synthetic typical farm situations were obtained by pooling and averaging the data on resource use, and input-output coefficients of 12 randomly selected farm situations from each size group. Labour requirements were estimated for various alternative cropping patterns based on different levels of technology. Budgeting technique was used to analyse the farm business.

The main source of farm power was bullocks on all the farms. The most prevalent source of irrigation was percolation wells on small farms and tubewells on medium and large farms.

An appraisal of existing labour use pattern

The existing labour use pattern of the small, medium and large synthetic farm situations is indicated in Appendices 1, 2 and 3.

The examination of these appendices shows that in case of small farms, there was overstocking of under-utilization of 154, 126, 127, 257, 80, 46 and 198 man-hours during the months of February, March, May, June, September, October and December respectively. Similarly, in case of medium farm, there were un-utilized man-hours varying from 13 to 198 during various months. In case of large farm organization, the under-utilization of man-hours were 117, 337, 118, 208, 140 and 76 during the months of February, March, April, June, September and December respectively. Thus, the overall under-utilization of labour was 8·85, 0·73 and 5·74 per cent on small, medium and large farm organizations respectively (Table 2). The operating capital also remained under-utilized on these farm situations to the extent of

Rs. 150, Rs. 453 and Rs. 350 on small, medium and large farm situations respectively.

Again, the existing per acre yield levels of almost all the crops were much below the potential and could be raised substantially through the application of improved agricultural methods and practices (Appendix 4).

Labour requirements with the adoption of yield-increasing technology

Farm adjustments to explore the possibilities of increassing farm incomes and to improve labour use on various synthetic farm situations, adopting different levels of technology, are indicated in Table 1.

(*a*) Labour use and farm income pattern on small size farm situation

This 10-acre farm situation used well irrigation and bullocks formed the main source of farm power.

Stage I : Human labour use based on improved crop plan and existing production techniques.

Recognizing that the small farmers can afford to make only marginal adjustments within the framework of the resource restrictions, alternative crop combinations were developed, using input-output coefficients of the average existing production techniques.

The introduction of new crop-combination resulted into an increased used of human labour the over-stocking of labour being reduced from 8·85 to 1·54 per cent (Table 2). In certain months such as May, June, and December, the labour surplus 127, 257 and 198 turned into a deficit of – 26, – 106, – 10 man-hours respectively (Appendix I). There was 5·95 per cent increase in capital requirement over the existing plan. The returns to fixed factors increased from Rs. 5825·97 to Rs 6971.44 (19·66 per cent) over the existing crop plan.

Stage II : Human labour use based on improved crop plan and improved production techniques.

In this alternative plan, shifts in crop acreage were incorporated by making use of improved production techniques and new crop varieties (Table 1) To implement this plan, human labour requirements increased by 33·15 per cent. The existing surplus labour (8·85 per cent) turned into a deficit of 21·36 per cent. The additional labour requirements were met by hiring 197, 240, 162, 343, 189 and 167 man-hours during the months of February, March, May. June, November and December respectively (Appendix I). Also, the farmer needed more working capital to implement the plan and borrowed Rs. 381 in the kharif and Rs. 228 in the rabi season. Thus the farmer needed 37·38 per cent additional working capital which was, however, within the borrowing capacity of the small farmers (Appendix I).

TABLE 2

Aggregate labour use pattern and seasonal operating capital requirements based on two levels of technology on different size *bullock operated synthetic farm situations*, Dehlon Development Block, I.A.D.P., District, Ludhiana.

Particulars	*Small Size Farm Situation (10 acres)*			*Medium Size Farm Situation (16 acres)*			*Large Size Farm Situation (22 acres)*		
	Man labour (Man hours)	*Woman labour (Woman hours)*	*Operatiug capital (total) (Rs.)*	*Man labour (Man hours)*	*Woman labour (Woman hours)*	*Operating capital (total) (Rs.)*	*Man labour (Man hours)*	*Woman labour (Woman hours)*	*Operating capital (total (Rs.)*
Total availability of the Resources	7648	1222	1629	7832	1440	3371	11792	1542	4363
Existing crop plan based on Av. Existing production technique.									
Requirements	6971	587	1479	7775	774	2918	1115	1355	4013
Surplus or deficit	+ 677	+ 35	+ 150	+ 57	+666	+ 453	+ 677	+ 187	+ 350
%age *surplus or deficit*	+ 8·85	+ 51·96	+ 9·21	+ 0·73	+46·25	+13·44	+ 5·74	+12·13	+ 8 02
Alternative I : Crop plan based on Av. Existing Production Techniques.									
Requirements	7530	800	1567	8798	1062	3198	11379	1874	4325
Surplus or deficit	+ 118	+422	+ 62	— 966	+ 378	+ 173	+ 413	— 332	+ 38
%age surplus or deficit	+1·54	+34·53	+ 3·80	— 12·33	+26·25	+ 5·13	+3·50	—21·53	+0·87
%age increase over the existing	8·01	36·28	5·95	13·16	37·21	9·59	2·37	38·30	7·77
Alternative II : Crop plan based on Improved Production Techniques									
Requirements	9282	760	2238	10543	1134	3919	14030	1752	5452
Surplus or deficit	—1634	+462	— 609	— 2711	+ 306	— 548	—2238	— 210	—1089
%age surplus or deficit	—21·36	+37 80	—37·38	—34·61	+21·25	—16·25	—18·98	—13·62	—24·96
%age increase over the existing	33·15	29·47	51·32	35·60	46·51	34·30	26·22	29·29	35·86

1. Monthwise labour requirements of each crop enterprise, based on two levels of technology, were first estimated to arrive at aggregate total requirements for the whole farm. Labour requirements for tending farm animals and other miscellaneous farm jobs. were added to each month's labour requirements before making final aggregation (Appendices 1, 2 and 3).

2. Assuming 26 days of 8 hours of each month during non-peak period and 28 days of 10 hours of each month during peak period. Man labour hours availability was worked out for different months.

3. Generally in the area, women attend to special jobs such as picking of cotton, removal of maize cobs and covers etc. So woman labour hours were considered separately.

TABLE 1

Cropping patterns based on different levels of technology on three Synthetic Farm Situations, Dehlon Block, I.A.D.P., Dist. Ludhiana

Crop Enterprises	*Small Modal Farm*			*Medium Modal Farm*			*Large Modal Farm*		
	Existing	*Alternative I**	*Alternative II*	*Existing*	*Alternative I*	*Alternative II*	*Existing*	*Alternative I*	*Alternative II*
Kharif Season (area in acres)									
Desi maize	2·92	1·00	0 50	4·07	1·00	0·75	5·18	1·50	2·50
Hybrid maize	0·28	2·00	2·00	0·53	4·00	4·25	0·64	4·00	4·50
American cotton	1·20	1·50	2·00	2·19	3·00	3·00	3·50	5·00	4·00
Desi cotton	0·64	1·00	0·50	0·44	0·25	0·50	0·83	1·35	1·50
Groundnut irrigated	0·77	1·25	2·00	0·69	2·50	3 00	2·77	3·00	4·50
Groundnut unirrigated	0·27	0·27	0·27	0·64	0·75	0 75	—	—	—
Sugarcane plant	0·25	0·25	0·50	0·31	0·50	0·50	0·70	1·00	1·00
Sugarcane ratoon	0·63	0·75	0·50	0·90	1·00	1·00	1·10	1·00	1·00
Fixed activity-kb. fodders	1·66	1.66	1·65	2·28	2·28	2·25	3·60	3 60	3·00
	8·62	9·68	10 00	12·05	15·28	16·00	18·32	20·45	22·00
Rabi Season (area in acres)									
Wheat after fallow	1·31	—	—	11·75	—	—	3·33	1·50	—
Wheat C-273 after kh. crops irrigated	2·73	5·25	4·48	3·64	8·40	7·00	6·19	9·50	9·50
Wheat C-286 after American cotton	1·20	1·50	2·00	2·19	3·00	3·00	3·50	5·00	4·00
Wheat PV-18 after kharif crops	—	—	1·00	—	—	2·00	—	—	2·75
Gram irrigated	0·42	—	—	1·33	—	—	1·54	—	1·00
Gram unirrigated	—	—	—	0·17	—	—	—	—	—
Barley	0·23	—	—	—	—	—	0·78	—	—
Wheat+Gram irrigated	0·73	—	—	—	—	—	—	—	—
Fixed activity-rabi fodders	1·86	1·86	1·25	2·34	2·34	1·75	3·63	3·63	2·75
Sub-total+Sugarcane	9·36	8·61	8·73	14·64	13·74	13·75	20·82	19·63	22·00
Total cropped area	17·98	18·29	18·73	26·49	29·02	29·75	39·14	40·08	44·00
Cropping intensity (%)	178·80	193·00	197·30	166·80	198·80	195·31	191·50	191·50	200·00
Total returns to fixed factors†	5825·97	6971·44	11620·86	9387·95	11289·49	18653·12	14585·59	16711·36	25778·02
percentage of increase over the existing income	—	19·66	99·46	—	20·25	98·69	—	14·57	76·73

* Alternative I =Existing production techniques+improved cropping plan.
Alternative II=Improved production techniques+improved cropping plan.

† Returns to fixed farm resources per acre were obtained, separately for existing and improved technology, by deducting variable costs from the gross returns per acre. Total returns to fixed factors were calculated by multiplying the returns acre by the number of acres sown.

Introduction of technological changes and shifts in crop combination, resulted in increasing returns to fixed resources from Rs. 5825·97 in the existing plan to Rs. 11620·86 in the alternative plan (Appendix 4).

(*b*) Labour use pattren on medium size 16-acre farm situation

Similar procedure was followed to examine the potentials of farm income on medium farms by introducing different types of yield increasing technology complementary to labour use.

It was found that introduction of yield increasing technology and intensive use of modern inputs resulted in substantial increase in labour requirements of this organization.

Stage I

The new crop combination based on existing technology demanded increased man-hours to the extent of 12·13 per cent. During the months of March, November and December, the surplus labour of 198, 147 and 107 turned to be a deficit of —36, —109, and —257 man-hours respectively (Appendix 2). Surplus working capital was reduced from 13.44 per cent to 5·13 per cent. This change resulted into an increase in farm income from Rs. 9387·95 to Rs. 11289·49 (20·25 per cent).

Stage II

To implement this plan, based on improved technology and better crop combinations, the farmer needed 2711 additional man-hours (34.61 per cent). Casual labour had to be hired in almost all the months varying from 62 to 502 man-hours. The woman-labour utilization increased by 46.51 per cent over the existing plan but there was no deficit. The aggregate increase in man-hours by incorporating the first and second stage technology was 13·16 and 35·60 per cent respectively (Table 2). Operating capital requirement increased by 16.25 per cent. The return to fixed factors increased from Rs. 9387·85 in the existing plan to Rs. 18653·12 (98·69 per cent) in the alternative plan.

(*c*) Labour use pattern on large size 22-acre farm situation

Similar analysis was repeated for large size synthetic farm situation.

Stage I

With the incorporation of crop shifts in the existing plan, there was marginal improvement in utilization of labour and the total surplus labour was reduced from 5·74 per cent to 2·37 per cent. The existing surplus of 187 woman-hours was turned into a deficit of 332 woman-hours. The returns to fixed factors increased from Rs. 14585·59 to Rs. 16711·36 (14·57 per cent) over the existing plan.

Stage II

To implement stage II plan, the farmer required 26·22 per cent man-hours and 29·29% woman-hours (Table 2). The surplus labour of 117, 337, 118, 208 and 76 man-hours turned into a deficit of 250, 263, 434, 526 and 216 man-hours during the months of February, March, April, June and December respectively. During the months of October, November and December, the cotton picking labour was hired to the extent of 195, 134 and 146 woman-hours respectively. Correspondingly, there was 35·86 per cent additional demand of operating capital and returns to fixed factors increased from Rs. 14585·59 to Rs. 25778·02 (76·73 per cent).

Comparison of labour use pattern and working capital requirements with two levels of technology

Stage I

Following the average existing production techniques, with the introduction of crop shifts the man-hours requirements increased by 8·01 per cent, 13·16 per cent and 2·37 per cent and woman-hours requirements increased by 36·28 per cent, 37·21 per cent and 38·30 per cent on small, medium and large farms respectively (Table 2). The increase in woman-hours was almost similar on all the farm size groups. The increase in man-labour requirements were more on medium farm as compared with the small and the large farm situation. This was due to greater irrigation facilities and higher cropping intensity on the medium farm organization. Also, the additional demand of working capital was more on medium farm than on small and large farms.

Stage II

By incorporating second stage technology, man-hours increased by 33·35, 35·60 and 26·22 per cent on small, medium and large farm situations respectively. The percentage increase in woman-labour was 29·47, 46·51 and 29·29 per cent on small, medium and large farms respectively. Only on large farms, woman labour-hours were in deficit by 13·62 per cent.

Requirement of operating capital increased by 51·32, 34·30 and 35·86 per cent on small, medium and large farm organizations respectively, but these were within the borrowing capacity of the farmers (Appendices 1, 2 and 3).

The reader may note that the present analysis does not include the more recent improvements in technology such as evolution of paddy T.N 1, Paddy IR-8 and hybrid bajra, etc. because of the different soil and climatic conditions of the study area. Wherever there is a possibility of introducing these varieties, the requirement of farm labour will be further increased. The cultivation of some of the latest high yielding varieties of wheat such as Kalyan 227, S-308, etc. also required additional labour hours, for the fine seed bed preparation, additional irrigations, additional care and harvesting and threshing operations. This means that if the cultivation of these latest varieties were incorporated in the farm plans, there will be a still further increase in labour use in all the farm organisations.

Conclusion

From this empirical study it could be concluded that Indian agriculture can and probably will move in the direction of more labour intensive production function for some time to come.

Special emphasis in agricultural development should, therefore, be laid on the introduction of such yield-increasing technology and modern inputs which are complementary to labour use in increasing farm incomes. This process could be followed up with profit till a stage is reached when Indian agriculture gets commercialized and necessitates capital-deepening rather than capital shallowing technology.

Appendix I

Monthly labour use pattern and seasonal capital requirements of small size bullock operated synthetic farm situation (10 acres), Dehlon Development Block, I.A.D.P. District Ludhiana.

Resources	*Unit*	*Availabilty o j the resource*	*Existing Cropping Pattern based on Existing techniques*		*Alternative crop plan based on existing techniques*		*Alternative crop plan based on improved production techniques*	
			Require-ments	*Surplus or deficit*	*Require-ments*	*Surplus or deficit*	*Require-ments*	*Surplus or deficit*
Man Labour :	Man hrs.							
January	,,	697	761	− 64	689	+ 8	775	− 78
February	,,	697	543	+ 154	548	+ 149	894	− 197
March	,,	518	392	+ 126	503	+ 15	758	− 240
April	,,	697	712	− 15	737	− 40	857	− 160
May	,,	697	570	+ 127	723	− 26	859	− 162
June	,,	697	440	+ 257	803	− 106	1040	− 343
July	,,	518	684	− 166	576	− 61	566	− 48
August	,,	518	574	− 56	458	+ 60	496	+ 22
September	,,	518	438	+ 80	358	+ 160	442	+ 76
October	,,	697	651	+ 46	651	+ 46	845	− 148
November	,,	697	707	− 10	774	− 77	886	− 189
December	,,	697	499	+ 198	707	− 10	864	− 167
TOTAL	,,	7648	6971	+ 677	7530	+ 118	9282	−1634
Woman labour :	Woman hrs.							
September	,,	305	51	+ 254	80	+ 225	40	+ 265
October	,,	306	197	+ 109	270	+ 36	260	+ 46
November	,,	305	135	+ 170	230	+ 75	240	+ 65
December	,,	306	204	+ 102	220	+ 86	220	+ 86
TOTAL	,,	1222	587	+ 635	800	+ 422	760	+ 462
Operating Capital :	Rupees							
Kharif cash	,,	823	755	+ 68	816	+ 7	1204	− 381
Rabi cash	,,	806	724	+ 82	751	+ 55	1034	− 224
TOTAL	,,	1629	1479	+ 150	1567	+ 62	2238	− 609
Credit Available :								
Kharif cash	,,	983	—	+ 983	—	+ 983	−481	+ 502
Rabi	,,	983	—	+ 983	—	+ 983	−227	+ 756
TOTAL	,,	1966	—	+1966	—	+1966	−708	+1258

Appendix U

Monthly labour use pattern and seasonal capital requirements on medium size bullock operated synthetic farm situation of 16 acres with tubewell as source of irrigation, Dehlon Block, I.A D.P. District Ludhiana.

Resources	*Unit*	*Availability of the resource*	*Existing Cropping Pattern based on Existing techniques*		*Alternative crop plan based on existing techniques*		*Optimum cropping pattern based on improved Production techniques*	
			Require-ments	*Surplus or deficit*	*Require-ments*	*Surplus or deficit*	*Require-ments*	*Surplus or deficit*
Man labour ;								
January	Man hrs.	714	948	— 234	746	— 32	955	— 241
February	,,	714	701	+ 13	805	— 91	995	— 281
March	,,	530	332	+ 198	566	— 36	632	— 102
April	,,	714	920	— 206	1066	— 352	1100	— 386
May	,,	714	627	+ 87	881	— 167	946	— 232
June	,,	714	553	+ 161	860	— 146	1216	— 502
July	,,	530	631	— 101	598	— 68	869	— 339
August	,,	530	433	+ 97	448	+ 82	592	— 62
September	,,	530	480	+ 50	227	+ 303	444	+ 86
October	,,	714	976	— 262	807	— 93	646	+ 68
November	,,	714	567	+ 147	823	— 109	1134	— 420
December	,,	714	607	+ 107	971	— 257	1014	— 300
TOTAL	,,	7832	7775	+ 57	8798	— 966	10543	—2711
Woman Labour :								
September	Woman hrs.	260	35	+ 325	20	+ 340	40	+ 320
October	,,	360	272	+ 88	330	+ 30	360	—
November	,,	360	254	+ 106	320	+ 40	340	+ 20
December	,,	360	213	+ 147	392	— 32	394	— 34
TOTAL	,,	1410	774	+ 666	1062	+ 378	1134	+ 306
Operating Capital :								
Kharif Cash	Rupees	1759	1461	+ 298	1695	+ 64	2194	— 435
Rabi Cash	,,	1612	1457	+ 155	1503	+ 109	1725	— 113
TOTAL	,,	3371	2918	+ 453	3198	+ 173	3919	— 548
Credit Available :								
Kharif	,,	1067	—	+1067	—	+1067	—425	+ 642
Rabi	,,	1067	—	+1067	—	+1067	—112	+ 955
TOTAL	,,	2134	—	+2134	—	+2134	—537	+1597

Appendix-IV (*contd*)

Practice/Enterprise	*Uuit*	*Groundnut irrigated*			*G. Nut unirrigated*			*Sugarcane Plan'*			*Sugar Ration*		
		Existing level	*Improv-ed level*	*Devia-tion +or—*	*Existing level*	*Improv-ed level*	*Devia-tion +or—*	*Existing level*	*Improv-ed level*	*Devia-tion +or—*	*Existing level*	*Improv-ed level*	*Devia-tion +or—*
A. **Output Pattern** (Yields)													
(*i*) Main Product	Qtls	5·00	9·00	+4·00	3·50	6·00	+2·50	16·00	31·00	+15·00	14·3	28·00	+13·7
(*ii*) By-product	,,	2·50	5·00	+25·0	2·00	3·00	+1·00	—	—	—	—	—	—
B. **Input Structure :**													
(1) Seed variety	—	Local	C501	—	Local	Punjab No. 1	—	Improv-ed	COJ39 COJ46 CO312	—	—	—	—
(2) Seed rate	Kgs.	25·00	26·00	+1·00	30·00	30·00	—	2800	3000	+200	—	—	—
(3) Seed treatment (Agroson/Agollal)	Gms.	—	75·00	+75·00	—	75·00	+75·00	—	Agollal	treat-ment	—	—	—
(4) Sowing time	—	June	June	—	With 1st shower of Monsoon rains.			March	Feb. 15 to March 15		—	—	—
(5) Interculture	No.	1	2	+1	—	1	+1	2	3	+1	2	3	+1
(6) Manure (F.Y.M.)	Tons.	1·62	—	—1·62	—	—	—	8·00	10·00	+2·00	10·0	8·00	—2·60
(7) Fertilizer :													
(*i*) CAN or Am. Sulphate	Qtls.	0·37	—	—0·37	—	—	—	1·57	2·50	+0·93	0·90	2·25	+1·35
(*ii*) Superhosphate	,,	—	1·00	+1·00	—	0·75	+0·75	0·20	1·00	+0·80	—	0·50	+0·50
(*iii*) Muriate of Potash		—	—	—	—	—	—	—	—	—	—	—	—
(8) Plant protection :													
(*i*) Sprays	No.	—	2	+2	—	2	+2	—	3	+3	—	3	+3
(*ii*) Endrine	Ltr.	—	—	—	—	—	—	—	1	+1	—	1	+1
(*iii*) Copper Oxy-chloride		—	Yes	—	—	Yes	—	—	—	—	—	—	—
(*iv*) BHC 10% dust	Kgs.	—	—	—	—	—	—	—	15	+15	—	—	—
(9) Irrigations	No.	2	3	+1	—	—	—	14	16	+2	14	16	+2
C. Returns to fixed farm resources	Rupees	336·94	656·15	319·21	237·84	409·71	171·87	568·61	1404·15	835·54	659·78	1476·19	816·41

Appendix III

Monthly labour use pattern and seasonal capital requirements on large size bullock operated synthetic farm situation of 22 acres with tubewell as source of irrigation Dehlon Block, IADP, District Ludhiana.

Resources	Unit	Availability of the resource	Existing Cropping Pattern based on Existing techniques		Alternative crop plan based on existing techniques		Optimum cropping pattern based on improved production techniques	
			Requirements	Surplus or Deficit	Requirements	Surplus or Deficit	Requirements	Surplus or Deficit
Man Labour :	Man hrs							
January	,,	1075	1120	− 45	1129	− 54	1153	− 78
February	,,	1075	1120	+117	1226	− 151	1325	− 250
March	,,	798	461	+337	508	+ 290	1061	− 263
April	,,	1075	957	+118	1031	+ 44	1509	− 434
May	,,	1075	1084	− 9	1247	− 172	1338	− 263
June	,,	1075	867	+208	1264	− 189	1601	− 526
July	,,	798	733	+ 65	883	− 85	1187	− 389
August	,,	798	780	+ 18	578	+ 220	805	− 7
September	,,	798	658	+140	384	+ 414	336	+ 462
October	,,	1075	1317	−242	900	+ 175	910	+ 165
November	,,	1075	1181	−106	1081	− 6	1514	− 439
December	,,	1075	999	+ 76	1148	− 73	1291	− 216
TOTAL	,,	11792	11115	+677	11379	+ 413	14030	−2238
Woman Labour :	Woman hrs.							
September	,,	385	66	+319	108	+ 277	120	+ 265
October	,,	385	449	− 64	662	− 277	580	− 195
November	,,	386	416	− 30	608	− 222	520	− 134
December	,,	386	424	− 38	496	− 110	532	− 146
TOTAL	,,	1542	1355	+187	1874	− 332	1752	− 210
Operating Capital :	Rupees							
Karif Cash	,,	2183	1897	+286	2106	+ 77	2809	− 626
Rabi cash	,,	2180	2116	+ 64	2219	− 39	2643	− 463
TOTAL	,,	4363	4013	+350	4325	+ 38	5452	−1089
Credit Available :								
Kharif	,,	1675	—	1675	—	+1675	−626	+1049
Rabi	,,	1675	—	1675	39	+1636	−464	+1211
TOTAL	,,	3350	—	3350	99	+3311	−1090	+2260

Appendix-IV *(concluded)*

Practice/Enterprise	*Unit*	*Gram irrigated*			*Rabi Fodder Berseem*		
		Existing level	*Improved level*	*Deviation +or—*	*Existing level*	*Improved level*	*Deviation +or—*
A. **Out-put Pattern** (Yields)							
(1) Main Product	Qtls.	3·56	7·00	+3·44	285	335	+50
(2) By-product	,,	3·56	7·00	+3·44	—	—	—
B. **In-put Structure :**							
(1) Seed Variety	—	Local	Pb. 7	—	Local	Mescavi	—
(2) Seed rate	Kgs.	18·00	16·50	—1·50	11·00	10·00	1·00
(3) Seed Treatment (Agroson/Solar heat treatment	gm.	—	—	—	—	Float out Khshni seeds	
(4) Sowing time	—	—	10th to 25th October		—	15th September to 15th October	
(5) Interculture	No.	—	—	—	—	Remove weeds	—
(6) Manure (F.Y.M.)	Tons	—	—	—	3·50	8·00	+4·50
(7) Fertilizers							
(*i*) CAN or Am. Sulphate	Qtls.	—	—	—	0·20	0·25	+0·05
(*ii*) Superphosphate	,,	—	1·00	+1·00	—	2·50	+2·50
(*iii*) Muriate of Potash	,,	—	—	—	—	—	—
(8) Plant Protection							
(*i*) Sprays	No.	—	—	—	—	—	—
(*ii*) Endrine	Ltrs.	—	—	—	—	—	—
(*iii*) Cu. Oxy-chloride		—	—	—	—	—	—
(*vi*) B.H.C. 10% dust	Kgs.	—	—	=	—	—	—
(9) Irrigations	No.	2	2	—	20	25	+5
C. Returns to fixed farm resources	Rupees	185·89	310·30	124·41	*Fed to cattle*		

Appendix IV

Comparative statement of the average existing production techniques, and improved production techniques, Dehlon Block, IADP, District Ludhiana, 1965-66.

Practice/Enterprise	Unit	Desi Maize			Hybrid Maize			American cotton			Desi cotton		
		Existing level	*Improved level*	*Deviation +or—*	*Existing level*	*Improved level*	*Deviation +or—*	*Existing level*	*Improved level*	*Deviation +or—*	*Existing level*	*Improved level*	*Deviation +or—*
A. **Out-put Pattern** (Yields)													
(*i*) Main Product	Qtls	7·85	12·00	+4·15	12·00	20·00	+8·00	3·26	5·50	+1·24	3·84	6·00	+2·16
(*ii*) By product	,,	9·50	13·50	+4·00	16·00	25·00	+9·00	8·00	8·00	—	8·00	8·00	—
B. **In-put Structure :**													
1. Seed Variety	—	Local	Selected	—	Ganga 101	Ganga 101	—	320F	320F	—	Local	231R	—
2. Seed Rate	Kgs	7·50	7·50	—	7·50	8·00	+0·50	5·00	7·50	+2·50	5·70	5·50	—0·20
3. Seed Treatment (Agroson/solar heat treated)	Gms	—	25	+25	—	25	+25	—	25	+25	—	—	—
4. Sowing time	—	July	July 5-22	—	July	July 1·15	—	April	April	—	May	May	—
5. Interculture	No.	2	3	+1	2	3	+1	2	3	+1.	1	1	—
6. Manures (F.Y.M.)	tons	8·78	6·00	—2·78	6·50	10·0	+3·50	6·20	—	—6·20	3·00	—	—3·00
7. Fertilizers													
(*i*) CAN or Am. Sulphate	Qtls	0·61	1·10	+0·51	1·50	2·25	+0·75	—	1·00	+1·00	—	—	—
(*ii*) Superphosphate	,,	—	0·90	+0 90	—	1·75	+1·75	—	0 50	+0·50	—	—	—
(*iii*) Muriate of Potash	,,	—	0·25	+0·25	—	0·50	+0·50	—	0·25	+0·25	—	—	—
8. Plant Protection													
(*i*) Sprays	No.	—	2	+2	—	3	+3	—	3	+3	—	—	—
(*ii*) Endrine	Ltr.	—	0·75	+0·75	—	1·25	+1·25	—	1·75	+1·75	—	—	—
(*iii*) Cu. Oxy-chloride		—	—	—	—	—	—	—	—	—	—	—	—
(*iv*) B.H.C. 10% dust	Kgs.	—	—	—	—	—	—	—	—	—	—	—	—
9. Irrigations	No.	5	5	—	4	6	+2	5	6	+1	4	4	—
C. Returns to fixed farm resources	Rupees	263·86	395·11	131·25	427·76	651·35	223·59	363·16	615·05	251·89	362·08	546·20	184·12

SELECTION LIMITS

by

B.R. MURTY*

The nature and magnitude of response to selection have been investigated by several workers of which the contributions of Fisher, Wright, Haldane, Robertson, Falconer and Thoday are of both theoretical interest and practical utility.

Panse (1957) has pointed out some situations where response to direct selection and selection by using discriminant function did not differ much and suggested the need for a critical evaluation on selection criteria and limits. A majority of the reports on selection limits have been predictions on the basis of analysis of population from crosses between highly selected group of varieties. In practically all of the cases, the predicted responses were mostly over-estimates and were never realised. Among the causes which have been considered to be responsible for limiting artificial selection are small sample size, correlated response, genotype-environment interactions, maternal effects and past history of selection. Among them, past history of selection appears to be a potent factor for influencing most of the responses to selection. In this paper, the work related to these aspects in crops belonging to diverse breeding systems, using *Brassica*, Linseed, Wheat, *Pennisetum* and *Sorghum* carried out during the past six years at I.A.R.I. will be outlined with reference to the theoretical predictions made by the above authors.

Material and Methods

The populations in Brassica belong to self-compatible and self-incompatible types representing the broad spectrum of variation of brown *sarson* in India. The linseed material comprised of diverse crosses involving elite Indian varieties as well as exotic types. In Pennisetum, the response has been studied in populations where dwarfing genes are introduced mainly to see the effect of a single gene substitution on correlated response in other traits including productivity. In the case of sorghum, the differential response of some crosses using the same male parent showing distinctly different types of gene action has been examined. In wheat, the effects of selection for developmental traits on limits to yield was considered. In all the cases, the plane of nutrition was maintained substantially high up to 100 or 120 kg. nitrogen per hectare so as to permit full expression of the genotype. In addition, a comparison of two methods of selection, namely, disruptive selection and directional selection were also investigated in Brassica.

---

* Division of Genetics, Indian Agricultural Research Institute, New Delhi.

Appendix-IV (*contd.*)

Practice/Enterprise	*Unit*	*Kharif Fodder*			*Wheat after Kharif*			*Wheat after Am. Cot.*			*Wheat PV-18 Irrign.*		
		Existing level	*Improv-ed level*	*Devia-tion +or—*	*Existing level*	*Improv-ed level*	*Devia-tion +or—*	*Existing level*	*Improv-ed level*	*Devia-tion +or—*	*Existing level*	*Improv-ed level*	*Devia-tion +or—*
A. **Out-put Pattern** (Yields)													
(*i*) Main Product	Qtls.	75·00	145·0	+70·00	6·65	10·06	+3·35	6·00	9·00	+3·00	—	22·00	
(*ii*) By-product	,,	—	—	—	12·11	14·00	+2·89	9·25	14·00	+4·75	—	22·00	
B. **Input Structure**													
(*i*) Seed Variety	—	Local	JS20 JS263	—	C273	C273	—	C273	C286	—	—	PV18	—
(*ii*) Seed Rate	Kgs.	26·00	23·00	—3·00	38·00	40·00	+2·00	40·00	40·00	—	—	35·00	
(*iii*) Seed Treatment (Agroson/Agollal /solar heat treated)	gms	—	—	—	—	Solar+Agroson treatment		—	Solar+Agroson treatment		—	Solar+Agroson treatment	
(*iv*) Sowing time	—	April to July		—	Nov.	Nov.		—	After 15th of Nov.		—	Nov.	—
(*v*) Interculture	No.	1	2	+1	1	2.4D spray	Spray 11 one		2-4D spray	Spray one	—	2-4D spray	—
(*vi*) Manures (F.Y.M.)	Tons	2·60	—	—2·60	—	—	—	—	—	—	—	—	—
(*vii*) Fertiliser :													
(*a*) CAN or Am. Sulphate	Qtls.	0·24	0·75	+0·51	0·87	1 00	+0·13	0·65	1·20	+0·55	—	3·00	
(*b*) Superphosphate	,,	—	—	—	0·20	0·50	+0·30	0·10	0·50	+0·40	—	1·50	
(*c*) Muriate of Potash	,,	—	—	—	—	0·20	+0·20	—	0·20	+0·20	—	0·50	
(*viii*) Plant protection													
(*a*) Sprays	No.	—	—	—	—	—	—	—	—	—	—	—	—
(*b*) Endrine	Ltr.	—	—	—	—	—	—	—	—	—	—	—	—
(*c*) Cu. Oxy-chloride		—	—	—	—	—	—	—	—	—	—	—	—
(*ix*) Irrigations	No.	3	5	2	5	5	—	4	4	—	—	6	—
C. **Returns** to fixed farm resources	Rupees	Fed to cattle			374·52	572·95	198·43	360·50	510·52	150·02	—	1152 81	—

Results

The selection of experiments in Brassica have clearly illustrated that the phenotypical uniformity of the material does not necessarily mean that there is limited genetic variability. The response to selection in this crop was considerable although mass selection, individual plant selection and modified mass pedigree method suggested were reported to be not successful. The selection results are compared for disruptive and two-way selection (Table 1).

Under disruptive selection for flowering time, five earliest and five latest plants in each row were selected. The progeny of the crosses between these earlies and lates belonging to each population were subjected to another cycle of selection and crossing as mentioned above, with negative assortative mating as outlined by Thoday (1958).

Release of Variability and *Asymmetry of Response* : The genetic gain was found to be substantial under disruptive selection for flowering time (Table 1). In the case of directional selection also, substantive divergence was obtained but the number of productive lines in that material was considerably less as compared to the material from disruptive selection experiments. It appears that selection for a vital character as flowering time has considerable correlated effects on others unless diversity and variability for them is maintained by *inter se* crossing of phenotypically alike individuals so as to release variability not previously available for selection due to tight linkage (McBride and Robertson 1963, Murty 1965). Moreover, there has been substantial correlated response for yield also and the diversity for seed productivity is also considerable. As a result, the proportion of selections which far exceeded the bulk were substantially more in the direction opposite to that in the previous history of selection. It was also found that the first generation of selection for a trait had such an effect that it was very difficult to revert to the level of the base population even by continued reverse selection subsequently. For example, if selection for earliness has been practised in the first generation, the mean could not be changed in the opposite direction to the original level even in the next three or four generations of intense selection for lateness. This was not necessarily true for initial selection towards lateness, indicating that the variability for lateness is to some extent restricted by conditions other than correlated response like genotype environment interactions. It would appear that the period of growth of Brassica is during a period with stress of moisture and temperature. Therefore, considerable genotype-environment interactions prevented the selection to be successful towards lateness whereas the early period of growth which is favourable had permitted the maximum advance for selection towards earliness. Under such conditions of asymmetry, any *a priori* prediction of correlated response would be incorrect. As pointed out by Bohren *et al* (1966), it is likely that asymmetry is more common and any symmetry found in any experiment would be a surprise. According to them, the most frequent contribution to asymmetry will be from loci contributing negetively to the covariance and having frequencies other than 0.5. Since it is very common that the gene frequencies are always other than 0.5, asymmetry of response is to be expected. Therefore, it can be concluded that predictions have to be based on additive genetic variance estimated in each generation thereby invalidating the existing theory for the prediction of correlated response. In any divergent selection for single character as was done in the case of Brassica, the effect of change in covariance will be $1/\sqrt{n}$ times as large as that of a model with one locus of each type, where n is the number of loci although the number loci of involved does not affect the presence or absence of a symmetry. Falconer (1960) and others concluded that changes in the basic parameters in the two directions may be responsible for the asymmetry of response while some considered differences in heritabilities in the two directions and in some cases genetic drift may be important, where genetic correlation is low.

From the results of Bohren *et al* (*loc cit*) it is clear that some mechanism other than genetic sampling and most probably the sensitivity of genetic covariance to changed gene frequency brought out by selection is responsible for asymmetric response. Actually such changes were observed both in *Brassica* and *Pennisetum* in our experiments.

Role of Recombination in Progress under Selection

Present data on Brassica has indicated that the variability released by disruptive selection was available for improving the means. This released variability was due to recombination. In the presence of linkage the direction of linkage disequilibrium generated by selection will determine the rate of change gene frequencies. Felsenstein (1965) showed that artificial selection on additive phenotype will generate negative linkage disequilibrium. Under those conditions, tight linkage will reduce the response to artificial selection. Moreover, he observed that there is a simple relationship between the sign of the epistatic parameters and the type of linkage disequilibrium generated by selection. It would, therefore, appear that the response to selection will depend on the breakdown linkage disequilibrium by recombination and the magnitude and the direction of the epistatic parameters. Therefore, it is to be expected that the response to artificial selection is limited in populations under directional selection as compared to that under disruptive selection. Another cause for the better response under disruptive selection lies in the superiority of full sib-mating and alternate parent off-spring-mating to other systems like selfing in the progress per generation towards greater homogeneity (Fisher, 1954; Gale, 1964). Therefore, disruptive selection in which there is considerable sib-mating appears to be more efficient than selfing. Full sub-mating is found to be superior to selfing in linseed, Pennisetum and sorghum also in this study.

The work of Thoday and Boam (1961) has clearly indicated that an accelerated response under selection in population which has plateaued previously, is mainly due to recombination. The results of Brassica confirm that recombination was the main cause for the accelerated response under disruptive selection since no new genes were introduced in the population during the experiment. Actually there was an increase in the chiasma frequency in the disruptive selection lines as compared to the base population (Harinarayana, unpublished). Therefore, the mating system as well changes in residual genotypic background might have changed both recombination frequencies and the recombination types and accelerated the response under selection.

Effect of Population Size on Limits to Selection

Another point of interest is that the results are at variance with the hypothesis of Lerner (1954) that outbreeders have heterozygote advantage and any directional selection reduces the fitness of the population. The roll of population size on response to selection appears to be very important than the magnitude of selection forces.

Even when selection is made for heterozygotes in small populations, the equilibrium gene frequency was found by Robertson (1952) as the most potent factor controlling the effect of such selection, since the effect of heterozygote advantage is at its maximum with central equilibrium frequencies close to 0·5 and least outside the limits 0·2 to 0·8. Thus the equilibrium gene frequencies rather than the actual magnitude of the selection forces control the rate of fixation and loss of genetic variation in small populations. The loss of genetic variation from a population may be very rapid if the equilibrium gene frequency is sufficiently extreme *i.e.*, beyond 0·8 or 0·2. Since in a majority of selection experiments the number of individuals chosen for next generation are few and the equilibrium gene frequencies are more often outside the central limits, the effect of small population size is considerable which can lead to fixation of a few alleles. This phenomenon might have been one of the causes of the limited advance realised under selection by previous workers in India in practically all cross-pollinated crops. While the expected total response gets gradually reduced with increase in the intensity of linkage and is appreciable at linkage values less than 0·10, the reduction in response to linkage is less pronounced for genes of smaller proportionate effect as compared to that for genes of fairly large effect. Even with close linkage of 10 map units, the reduction in response due to linkage is found to be of a minor order in comparison to the reduction in response to restricted population size alone, in the computer simulation studies by Latter (1965). Thus, the poor selection response in practically all crops during the past three decades in India can be mainly traced to restricted size of the effective breeding population.

Developmental Traits and Selection Limit

In the case of sorghum, a comparison of the nature of gene action in two heterotic hybrids CSH 1 and CSH 2 having a common female parent

MSCK 60 has brought out that diverse genetic mechanisms were involved in the same phenotypic end product. Selection in segregating populations in both the hybrids resulted in the isolation of some lines superior and at least equal to the F_1 in the cross CSH 1 and little success in the cross *CSH* 2. Moreover, the association of yield with flowering time which considerably influences the yield has revealed that considerable physiological association rather than linkage was responsible for the limited response to selection in CSH 2 as compared to the substantial additive gene action for flowering time and yield in CSH 1 with linkage rather than physiological association responsible for the association between yield and maturity.

The components of heterosis were estimated from generation means as outlined by Hayman (1958), where *d, h, i, j* and *l* are additive, dominance, additive × additive, additive × dominance and dominance × dominance effects respectively. Moreover, *d* and *i* components are important in heterosis of CSH 1 while *h* and *l* are important in CSH 2 (Rao, 1967 unpublished thesis). Thus, it would appear that divergent mechanisms of heterosis in these two hybrids are associated with different developmental mechanisms.

In the case of linseed the cause for earlier reports of limited success is found to be due to selection in the early segregating generations for a single trait, such as rust resistance. However, when selection for developmental features such as synchronisation of branching was carried in BIPs of the same cross combination, the extent of recombination obtained due to the released variability is substantial to get responses to selection not achieved earlier (Anand 1967). The yield of such new selection are given in Table 2. Moreover, population performance rather than individual performance is considered for selection (Murty, Arunachalam and Anand, 1967 ; Anand, 1967, Robertson 1962).

In the case of wheat, evidence has been obtained (Roy and Murty, 1967), that the environment in which selection is made is also important for the degree of response selection. Selection is more favourable environment for developmental traits like synchrony of tillering has permitted improvement in yield and wide adaptation as compared to selection under stress environment. Among the selection which did well under high, medium and low fertility and rainfed conditions from the same set of crosses, 86% of lines are those selected for synchrony of tillering under high fertility.

In the case of Pennisetum, the influence of dwarfing genes was found to be considerable in changing the character associations to a substantial degree (Murty and Tiwari, 1967). The diversity among the derivatives is so high that it cannot be accounted by the effect of only the loci for dwarfing. It is likely that introduction of new alleles of diverse origin and selection for dwarfing and associated traits must have increased the variability. The activation of some dormant loci, thereby increasing the variation available for selection cannot be ruled out. Moreover, the expression of an allele can be influenced by the residual genetic background, thereby contributing to changes in genetic variation. The loci of dwarfing are found to be of complex nature and their introduction into a material which has previously not responded to selection might bring in alleles at several other associated loci which are divergent and therefore generate new variability which would promote the response to selection. In these cases also sib-mating *inter se* in the early generations for plant type and population performance rather than single plant yield has substantially increased the yields as compared to the derivatives from crosses between stable lines of advanced generations.

Role of Dominance and Heterozygote Advantage in Selection Limit

A population can reach a selection limit while still retaining genetic variation due to continuous selection for heterozygote. Under mass selection, Kojima (1961) compared the responses to selection in the presence of dominance in finite populations and found that the joint effects of dominance and finite size of population could cause a considerable bias in the usual prediction equation under mass selection. Under a given gene frequency and intensity of selection, the variance of change in gene frequencies is a function of additive and dominance components only. However, in the absence of dominance effects of genes the expected gains under mass selection were not seriously different with changes in the size of the population.

Effect of Selection Intensity and Previous History of Selection on Response to Selection

To get the maximum possible advance from a given initial population it is necessary to start with a high value of *Ni* where *N* is population size and *i* is intensity of selection so as to fix with high probability all the rare but desirable genes. The choice of *Ni*, therefore, will depend upon the previous history of the population both in terms of its selection history and its population size. The more highly selected a strain is the smaller is the number needed for keeping it. Therefore, greater care should be given in an unselected strain so that the desirable alleles which are likely to be at low frequency are not lost by genetic drift. Moreover, for a gene with selective advantages, the change of fixation is a function only of *Ns* where *N* is the effective population size. Robertson (1960, 1961) has shown from computer studies that if we make populations by crossing selected lines, and then use a proportionately higher value of *Ns* we have the same expected limit as if we had used the higher value all the time. Moreover, genetically superior individuals will tend to be mostly inbred under selection not only at loci carrying genes determining the character in question but at all loci. Robertson obtained evidence that inbreeding due to selection for outstanding individuals may arise for several generations after their use *i.e.* there is a time lag in the effect. Inbreeding under individual selection will tend to be greater than that calculated from the actual number of parents when heritability of the character and intensity of selection are high. The most interesting observations by Robertson (1960) on limits in artificial selection are that the effective population size in any selection programme is dependent on the intensity of selection, and that linkage too plays an important role on the magnitude of response. While one would expect that the greater the number of generation over which selection differential is spread, greater will be the response because of higher probability of recombination during this period, this does not appear to be true in several cases. Ultimately, a family selection or progeny testing programme always involves some sacrifice of ultimate response for the sake of greater immediate gain. Therefore, it is necessary not to intensely select in the early generations.

Breeding System and Selection Advance

The results reported on the response to selection in crops with diverse breeding systems have revealed that there is no single factor or a set of factors which may be associated with the breeding structure for limiting progress in either self or cross-pollinated crops. It is also clear that in a majority of the cases, irrespective of the breeding system, the small sample size must have resulted in a substantial correlated response and loss of genetic variability due to genetic drift, since most of the productive alleles are in low frequencies in any population which has undergone very limited selection. Even in crosses between varieties which are highly selected, there is considerable difficulty in changing the constellations of genes associated with previous selection. In such cases breaking of the linkage by mating the individuals in the early generations *inter se* and maintenance of genetic heterogeneity and phenotypic uniformity for developmental traits rather than yield itself appears to be important.

The advances made under selection in populations which are found to have plateaued, also have revealed interesting information. There is considerable variability among members within a population less subject to human selection, which, if released by crossing the members *inter se* will be available for selection. The introduction of new genes from divergent resources has a phenomenal effect of improving the productivity of materials which have failed to respond to selection. Such accelerated response is mainly due to recombination as substantiated by the selection experiments by Thoday and Boam (1961). White maternal effects could considerably result in failure to obtain predicted advance from models which did not take these effects into consideration, data are available that these maternal effects can be considered as functions of the additive genetic variation and the mean or the parental genotype. Therefore, selection for high combining ability or for a system of genes which improve the adaptive organisation of a population will ensure the minimising of the role of the maternal effects in the advance under selection.

While it is true that discriminant function used by earlier workers has been useful in some cases, it was a failure in several instances mainly due to inadequacy of diversity in the material and the inadequacy of the models employed. However, the choice of parents using multivariate analysis for assessing gentic divergence selection for developmental traits in the early generations, by evaluation of population performance, maintenance of genetic heterogeneity with phenotypic uniformity appear to help in accelerating the advance under selection and the magnitude of response. It is clear that limits to artificial selection in the present day cultivated varieties are not yet reached. It is likely that selection for physiological traits such as efficient transfer of energy into grain or the consumable end product as in sorghum and choice of genotypes that efficiently utilise the nutrients in the early stages of growth under stress environment will be vital. The present barriers to selection do not appear to be as formidable as outlined in the early years of selection in this country in different crop plants.

REFERENCES

Anand, I. J. (1967). Genetic diversity and gene interaction for some components of yield in linseed (*Linum usitatissimum. L*). Ph. D. Thesis, University of Agra.

Bohren, B.B., W.G. Hill and A. Robertson (1966). Some observations on asymmetrical correlated responses to selection. *Genet. Res.*, **7** : 44-57.

Felsenstein, J. (1965). The effect of linkage on directional selection. *Genetics*, **52**: 349-363.

Fisher, R.A. (1959). An algebraic examination of junction formation and transmission in parent-offspring inbreeding. *Heredity*, **13** : 179-186.

Gale, J.S. (1964). Some applications of the theory of junctions. *Biometrics* **20** : 85-117.

Hayman, B.I. (1958). The separation of epistatic from additive and dominance variation in generation means. *Heredity* **12** : 371-391.

Kojima, K. (1961). Effects of dominance and size of population on response to mass selection. *Genet. Res.*, **2** : 177-188.

Latter, B.D.H. (1965). The response to artificial selection due to autosomal genes of large effect. II. The effects of linkage on limits to selection in finite populations. *Austr. J. Biol. Sci.*, **18** : 1009-1023.

McBride, G. and A. Robertson (1963). Selection using assortative mating in *Drosophila melanogaster*, *Genet. Res.*, **4** : 356-369.

Murty, B.R. (1965). The effect of disruptive selection on self incompatibility in *Brassica campestris* var. brown *Sarson- Proc. Symp. on the Mutational Processes*. Prague, pp. 105-107.

Murty, B.R. and Arunachalam V. (1966). The nature of divergence in relation to breeding system in some crop plants. *Ind. J. Genet.* **26A** : 188-198.

Murty, B.R., Arunachalam V. and Anand I J. (1967) Diallel and partial diallel analysis of some yield factors in *Linum usitatissimum*. *Heredity* **22** : 35-42.

Murty, B.R. and Tiwari J.L. (1967). The effect of dwarfing genes on gentic diversity in *Pennisetum typhoides*. *Ind. J. Genet.*, 28 : (in press).

Panse, V.G. (1957) Genetic of quantitative characters in relation to plant breeding. *Ind. J. Genet. & Pl. Br.*, **17** : 318-328.

Rao, C.R. (1960). Multivariate analysis– an indispensable statistical aid in applied research. *Sankhya*, **22** : 317-338.

Rao, N.G.P. (1967). Studies in heterosis in sorghum. Ph. D. Thesis (unpublished). Bihar University 1967.

Rao, N.G.P. & Murty B. R. (1963). Growth analysis in grain sorghum of Deccan. *Ind. J. Agri. Sci.*, **33** : 155-162.

Robertson, A (1960). A theory of limits in artificial selection. *Proc. Roy. Soc. Series B*, **951** : 234-249.

Robertson, A (1961). Inbreeding in artificial selection programmes. *Genet. Res.*, **2** : 189-194.

Robertson, A. (1962). Selection for heterozygotes in small populations. Genetics. **47** : 1291-1300.

Roy, N.N. and Murty B.R. (1967). Response to selection for wide adaptation in bread wheat. *Curr. Sci.*, **36** : 481-482.

Thoday, J.M., and T.B. Boam (1961). Regular responses to selection. I. Description of responses. *Genet. Res.* **2** : 161-176.

Thoday, J. M., (1958). Effect of disruptive selection. The experimental production of a polymorphic population. *Nature,* **181** : 1124-1125.

TABLE 1

Response to selection under disruptive selection for flowering time

Population	50 % Flowering time in days			Seed yield in gms/plot		
	Base Population	Earliest	Latest	Base	Early	Late
1. K. Lotni 17	60·0	49·5	78·5	124·8	227·2	248·9
2. K. Lotni 27	60·5	48·5	72·0	175·5	330·2	273·2
3. K. Tora 5905	49·5	48·5	74·5	177·3	236 5	260·0
4. K. Tora 5907	75·5	57·0	82·0	238·0	335·9	393·9
5. GBST	61·5	54·0	66·0	262·2	347 6	398·4

Response under two-way selection for flowering time

	Days to flowering			Yield in gms./plot		
	Base popn.	Early	Late	Base popn.	Early	Late
1. K. Tora 5907	70	40	80	126	208	168
2. IARI 117	76	60	89	168	170	204
3. Assam Local	42	38	68	177	235	216
4. Pusa BST 2	59	40	83	94	215	175
5. GBS 1	47	40	70	98	264	300

TABLE 2

Selection response for yield and maturity in crosses in some crop plants

Parents	Brown Sarson[1]		Linseed[2]			Sorghum[3]			Pennisetum[3]			Wheat[3]		
	A	B	Parents	A	B	Parents	A	B	Parents	A	B	Parents	A	B
Kanpur Lotni 5907	76	238 gms	M10	69	850 kg	IS34	66	37 gms	D_2	64	1319 kg	E 4871	91	471 gms
—	—	—	RR9	81	906 kg	CK60	61	35 gms	IP81	53	1148 kg	Sonora 64	89	477 gms
Selections (*a*) obtained	57	336 gms	Selection$_1$	72	1452 kg	Selection$_1$	63	45 gms	D174	63	2500 kg	Selection$_1$	89	646 gms
(*b*)	82	394 gms	selection$_2$	83	1497 kg	Selection$_2$	67	36 gms	—	—	—	Selections$_2$	89	562 gms
Percentage increase of selections over best parent	—	153%	—	–	163%	—	–	109%	—	—	190%	–	–	127%

A=Days to 50 % flowering.
B=Yield in gm/unit or Kg/hectare.
1. Disruptive selection.
2. Directional Selection in BIP's
3. Selection in F_2 for developmental traits and selection for yield on population basis.

Largest incidence was observed in *Sahiwal* and *Sindhi* and the least in *Tharparkar*. The differences amongs breeds were not only highly significant, but even within a given breed, where such an analysis was possible, as in *Hariana* and *Sahiwal*, highly significant variation was observed. There was also significant difference in incidence between the two sexes, the females showing higher incidence than the males.

Depending upon the location of the SNTs on the udder, 3 types of SNT could be recognised. They were (*a*) intercalary, (*b*) caudal and (*c*) ramal. The most frequent were the intercalaries followed in order by caudals and ramals. Ramals were totally absent in the samples of animals studied in *Gangateri*, *Sindhi* and *Tharparkar* breeds. The rates of incidence of the three types of SNTs in the 5 breeds were as under :

Breed	*Intercalary* %	*Caudal* %	*Ramal* %
Gangateri	13·85	2·16	—
Hariana	8·57	3·22	0·69
Sahiwal	19·40	7·46	1·11
Sindhi	29·23	1·53	—
Tharparkar	9·39	2·20	—

Further, in case of *Hariana* and *Sahiwal*, farm to farm analysis of variation was possible and showed that the intercalaries varied from farm to farm also. Significant variation between sexes in case of both intercalaries and caudals was present, the incidence being higher in females than in males (Singh and Prabhu, 1966*a*).

Genetic analysis carried out on *Hariana* and *Sahiwal* data established the fact that highly significant effect due to sire was found in the former and not in the latter. The half-sib correlations was high (0·32 ± 0 0024) in the former and small and negligible (0·048 ± 08·0200) in the latter breed of cattle (Singh and Prabhu, 1966*b*).

2. Horn Pattern and Size

The work was conducted on *Hariana* and *Tharparkar* breeds only. Three types of horns were (*a*) tight, (*b*) loose and (*c*) mixed *i.e.* one tight and another loose. The per cent incidences of these 3 types in the two breeds were as follows :

Breed	*Tight*	*Loose*	*Mixed*
Hariana	72·77	23·92	3·31
Tharparkar	91·58	7·37	1·05

In Table 1 are presented data regarding the trait studied, breeds considered, the number of farms per breed visited and the total number of animals in each farm.

TABLE 1

Details of experimental material

Sl. No.	*Trait*	*Breeds*	*No. of farms*	*Total No. of animals*
1.	Supernumerary teats (SNT)	**Gangateri**	1	231
		Hariana	7	3872
		Sahiwal	3	804
		Sindhi	1	65
		Tharparkar	1	181
2.	Horn patterns and size	**Hariana**	5	694
		Tharparkar	1	95
3.	Skin thickness	**Hariana**	4	473
		Sahiwal	4	334
		Sindhi	2	121
		Tharparkar	2	222
4.	Head size and shape	**Hariana**	5	660
		Sahiwal	4	283
		Sindhi	2	81
		Tharparkar	2	270
5.	Udder size and shape	**Hariana**	4	193
		Tharparkar	2	148
6.	Teat size and shape	**Hariana**	4	193
		Tharparkar	2	148

Observations on item No. 1 was carried out by Singh (1965) ; on item 2 by Narayan (1967) ; on item 3 by Bhatia (1967) and Mohan (1967) ; on item 4 by Chatterjee (1967) and Choudhury (1967) and on items 5 and 6 by Sharma (1967).

For estimating heritability, the method described by Hazel and Terril (1945), namely the half-sib correlation method was adopted. For other statistical treatment of the basic data, the standard procedures described in Snedecor (1956) were followed :

Results

I. SNT in Indian Cattle

An overall rate of 14.68% was observed for Indian cattle in the 5 breeds studied. The rates were as follows for the different breeds.

Gangateri	...	15·59%
Hariana	...	12·26%
Sahiwal	...	25·12%
Sindhi	...	30·77%
Tharparkar	...	11·60%

Variation in mean horn length in the two breeds according to age is summarised in Table 3.

TABLE 3

Variation in horn length and circumference at base (cms.)

Age in years	Length in cms.		Circumference in cms.	
	Hariana	*Tharparkar*	*Hariana*	*Tharparkar*
1–2	4·07	2·51	12·46	11·39
2—3	5·92	5·87	14·49	14·60
3—4	9·76	10·83	17·07	17·58
4—5	11·36	12·52	16·61	17·65
5—6	13·27	17·20	16·50	19·52
6—7	15·80	21·56	18·28	20·43
7—8	17·68	18·61	18·14	18·54
8—9	20·41	21·44	17·91	19·19
9—10	20·71	25·56	19·17	21·06
10—11	20·62	—	19·24	—
11—12	20·25	25·34	17·84	20·42

In *Hariana* variation among age groups as well as among farms was significant. The interaction, farms × age group was significant showing that horn length in age-groups had varied with farms. In *Tharparkar* variation among age-groups was significant. That among farms could not be calculated as there was only one farm considered. There was significant difference in horn length between breeds and age groups between breeds.

Results similar to that of horn length were also found in horn circumference measurements.

In *Hariana* heritability for horn length varied from 27% to 55% and for horn circumference from 11% to 41%. In *Tharparkar* the respective figures for horn length and circumference were 11% and 17% (Narayan, 1967).

STUDIES ON VARIATION IN CERTAIN PHYSICAL TRAITS OF SOME INDIAN CATTLE BREEDS

By

S.S. PRABHU*

Systematic analysis of any variation of a trait present in domesticated animals on this country was done for the first time, perhaps by Sukhatme (1944) when he reported on the working of the goat breeding scheme at Etah. Since then we have a number of reports on production characteristics largely based on accumulated data at cattle breeding farms; notable amongst these are those of Amble and his co-workers (1958*a*, *b*, 1960, 1963). No attention appears to have been given to the actual observation of physical traits and their variation as present in the different cattle breeds of India. According to published reports (Randhawa, 1962) there are 26 different cattle breeds which naturally fall under 3 categories namely, milch, general utility and draught breeds. Olver (1938) had grouped them into 5 divisions, depending upon their general appearance, size, form and function. Ware (1942) and Phillips (1934) helped in crystallising the scheme of classification suggested by Olver.

Some interesting data regarding the variation in respect of certain characteristics as seen in selected types of the breeds are presented in ICAR bulletins reissued as a single bulletin (ICAR, 1960) and that brought out by the FAO (Phillips and Joshi, 1953). The data, however, were small samples, from 'selected' farms and animals. They could not be taken as representing the actual variation present. Further, no figures were available regarding the nature of these variations. To fill in this gap in our knowledge about the extent and nature of variations of physical traits present in Indian cattle breeds, a systematic studywas takenin hand by the Animal Genetics Division of the Indian Veterinary Research Institute, Izatnagar, U.P. with the assistance of students supplicating for the Master's degree in Veterinary science. In the present paper are summarised the main findings of these studies.

Meterial and Method

The procedure followed was to select a number of farms of a given breed, where sufficient number of animals of a given breed were located. In case of some traits all the animals found in the farm were considered, while in others only a selected number were studied. The selection was so done that for a given sire sufficient number of daughters were available.

---

*Indian Veterinary Research Institute.

In a subsequent study, however, it was discovered that measurements made by Vernier calipers of the ordinary type were subject to error and depended upon the pressure applied. This defect could be overcome by use of a German Vernier Calipers which could be adjusted to give uniform pressure. Using these calipers, different breeds and samples, a more comprehensive study on the skin thickness in Indian cattle breeds was conducted (Bhatia, 1967 ; Mohan, 1967).

Overall mean skin thickness found in the different cattle breeds is summarised in Table 5.

TABLE 5

Overall mean skin thickness (mm)

Age group (years)	*Hariana*	*Sahiwal*	*Sindhi*	*Tharparkar*
0–1	4·135	5·379	5·198	4·271
1–2	4·899	7·035	7 521	4·908
2–3	5·622	7·671	7·932	—
3–4	5·596	8·394	8·256	6·494
4–5	5·623	8·425	8·414	6·619
5–6	5·936	8·610	8·196	6·686
6–7	5·916	8·521	7·912	6·554
7–8	5·703	8·441	—	6·642
8–9	5·492	8·102	—	6·386
9–10	5·725	8·030	7·290	6·552
10 and above	5·574	—	—	6 349

Results showed that the breeds *Sahiwal* and *Sindhi* had thicker skin than the *Hariana* and *Tharparkar*. The least skin thickness was found in *Hariana*. *Tharparkar* was in between *Hariana* and *Sahiwal* and *Sindhi*. Significant variation due to both age and site of measurement was found in all breeds. In all, except *Sindhi*, where it occurred between 2 and 3 years age, skin thickness got stabilised between the ages of 3 and 4 years. Thickest skin was found on the back and thinnest in the elbow region. In general, skin thickness increased as one went from ventral to dorsal region and from anterior to posterior region. Variation due to farms which was

There was significant difference from farm to farm in respect of the incidence of tight and loose horns. At Bassi farm for example, 93·18% were tight and 4·54% loose, while at Dumraon farm the respective figures were 55·68% and 38·64% and at Izatnagar 48·00% and 47·00%. This showed that the horn type has not been stablished in *Hariana* as yet, while in *Tharparkar*, it had been, 91·58% of the horns were tight and only 7·37% were loose.

Horn pattern has been described as horn direction throughout its length in comparison to the body axis, the head, neck and body for this purpose being taken as remaining in a straight line, Ten "patterns" of horn were recognised. Their description and the rate of incidence in the two breeds as found by Narayan (1967) are given in Table 2.

TABLE 2

Incidence of horn patterns in Hariana and Tharparkar

Sl. No.	*Description of pattern*	*Incidence %*	
		Hariana	*Tharparkar*
1.	Lateral	4·07	—
2.	Lateral-upward	33·49	8·99
3.	Lateral-upward-inward	33·33	80·89
4.	Lateral-upward-forward	8·61	2·26
5.	Lateral-upward-backward	6·42	—
6.	Lateral-forward	0·94	—
7.	Lateral-downward	10·80	4·49
8.	Lateral-downward-backward	1·09	—
9.	Lateral-downward-forward-upward	0·16	—
10.	Lateral-upward-inward-backward	1·09	3·37

Majority of the horns in *Hariana* were of lateral-upward and lateral-upward-inward types, while in *Tharparkar* it was of lateral-upward-inward type only. While 10 types of patterns were noted in *Hariana*, only five types of patterns were found in *Tharparkar*, showing that there was considerable admixture of the original breed, with other breeds. This was corroborated by significant variation in the incidence of occurrence of the patterns among farms of *Hariana* breed.

TABLE 8

Overall mean orbital width (cm)

Age group (years)	*Hariana*	*Sahiwal*	*Sindhi*	*Tharparkar*
1—2	21·60	27·50	29·66	22·53
2—3	25·79	29·89	31·91	26·85
3—4	26·91	30·88	32·10	27·64
4—5	29·00	31·54	31·20	28·39
5—6	30·16	32·24	32·57	31·44
6—7	30·99	31·95	34·51	31·72
7—8	30·86	33·58	—	31·78
8—9	31·16	32·99	—	3.-12
9—10	31·77	—	—	31·87

TABLE 9

Overall mean maxillary width (cm)

Age group (years)	*Hariana*	*Sahiwal*	*Sindhi*	*Tharparkar*
1—2	17·16	21·34	23·07	18·34
2—3	20·62	24·85	24·48	22·75
3—4	24·72	25·80	25·49	24·72
4—5	26·82	26·27	25·51	27·45
5—6	29·59	26·84	26·03	29·55
6—7	29·38	27·22	27·25	29·81
7—8	29·62	27·40	—	29·74
8—9	30·36	26·93	—	30·50
9—10	31·01	—	—	31·37

TABLE 10

Overall mean rams length (cm)

Age group (years)	*Hariana*	*Sahiwal*	*Sindhi*	*Tharparkar*
1—2	21·36	26,48	28·46	22·28
2—3	28·62	30·01	30·43	29·68
3—4	32·39	31·06	30·88	33·34
4—5	35·62	31·99	30·86	36·96
5—6	39·57	32·23	31·10	39·32
6—7	40·36	32·01	31·53	40·64
7—8	41·03	31·93	—	41·71
8—9	42·30	30·77	—	42·22
9—10	43·12	—	—	42·61

3. Skin thickness

The earliest study on the skin thickness in Indian cattle breeds was done by Chet Ram (1955, 1968). *Hariana* breed was used for the study and all observations were taken from animals belonging to a single farm (Izatnagar). 404 animals of different age groups were included in the study, which not only included animals of both sexes, but also of both growing and adult ones. Seven regions were selected namely, (*a*), throat, (*b*) dewlap, (*c*) neck, (*d*) elbow flap, (*e*) navel (*linea alba* anterior to the navel), (*f*) flank and (*g*) back. Measurement was done of folded skin with the help of a Vernier calipers. In Table 4 are summarized the data collected by him (Chet Ram, 1968).

TABLE 4

Mean skin thickness in Hariana (mm)

Age	*Mean overall skin thickness*	
	Males	*Females*
Months		
0—3	3·76	3·40
3—6	3·31	3·56
6—12	3·83	3·76
Years		
1—2	5·22	3·90
2–3	6·10	4 37
3—4	6·71	5·67
4—5	8·72	6·44
5–6	8·60	6·15
6–7	7·27	6·31
7–8	8·28	6·22
8–9	8·91	6·24
1–15	8·51	6·73

Significant variation was found between age groups, regions and sex. In addition, age group × region, age group × sex and age group × region × sex interactions were also found to be significant. Females had thinner skin thickness than males. Upto one year, the sexes did not differ significantly. Navel and dewlap skin thickness approximated to the mean overall average of all the 7 regions studied.

5. Udder shape and size

The work was conducted only on *Hariana* and *Tharparkar* breeds of cows. Three types of shapes were recognised. They were (*a*) bowl, (b) round and (*c*) goat type udders. The incidence of these three types in the two breeds was as follows :

Breed	*Bowl* %	*Round* %	*Goat* %
Hariana	93·3	2·6	4·1
Tharparkar	89·6	2·0	8·1

In both the breeds, the bowl shaped was the most common and round type, the least. The incidence of goat type udders was twice as frequent in *Tharparkars* than in *Harianas*. The difference was statistically significant.

Following measurements were made by udder size.

(*a*) **Udder length**—Length from rear attachment to front of udder where the fore-udder blends smoothly with the body.

(*b*) **Udder width**—Distance between two lateral lines of attachments of the udder to abdominal wall beneath the flank.

(*c*) **Udder depth**—Difference between lengths of barn floor to base of udder and from barn floor to the lowest point of udder floor.

The actual measurements taken are as given in Fig. 2.

Data collected by Sharma (1967) are summarised in Table 12.

The figures are overall means of all farms.

TABLE 12

Overall mean udder measurements in Hariana and Tharparkar cows (cms)

Lacta-tion order	*Hariana*			*Tharparkar*		
		Udder			*Udder*	
	Length	*Width*	*Depth*	*Length*	*width*	*Depth*
1.	40·96	48·67	17·55	43·57	55·84	22·13
2.	43·09	49·96	17·95	51 09	62·71	25·29
3·	45·78	52 83	19·59	51·35	63·32	26·21
4·	46·64	53·12	19·64	53·31	63·53	25·91
5·	45·38	55·94	21·65	54·44	65·36	26·80
6·	46·45	55·02	23·37	54·74	64·41	28·41
7·	49·15	54·75	23·18	62·10	63·70	25.95
8·	—	—	—	53·33	64.77	29·20
9·	—	—	—	52·35	62·25	26·90

significant was also noticed in respect of animals of a given breed. Heritability estimates of skin thickness on the basis of overall averages for the different farms and breeds are summarised in Table 6.

TABLE 6

Heritability estimates of skin thickness

Farms	*Hariana*	*Sahiwal*	*Sindhi*	*Tharparkar*
1	0·08±0·14	0·04±0·06	0·37±0·26	0·76±0·20
2	0·20±0·22	0·29±0·23	0·43±0·38	0
3	0·14±0·14	0·80±0·69	—	—
4	—	0·05±0·07	—	—

Only in respect of *Tharparkar* high estimate was obtained. In rest of the cases, on the whole, the estimates were low, and when high, had low precision.

4. Head size and shape

Following measurements were made on the head of *Hariana, Sahiwal, Sindhi* and *Tharparkar* breeds of Indian cattle.

(*a*) *Head length*—Distance between the frontal eminence and terminal end of upper commissure of lip (Fig. 1.).

(*b*) *Orbital width*—Distance between the two outer canthus of eyes (Fig. 1).

(*c*) *Maxillary width*—Distance between the outer margin of the two maxillary tubersities (Fig. 1).

(*d*) *Length of ramus*—Distance between angle of mandible and symphysis mandibulae (Fig. 1).

Data collected by Chatterjee (1967) and Choudhury (1967) are summarised in Tables 7-10.

TABLE 7

Overall mean head length (cm)

Age group (years)	*Hariana*	*Sahiwal*	*Sindhi*	*Tharparkar*
1—2	39·06	45·50	50·87	38·49
2—3	41·75	51·87	53·21	43·82
3—4	51·75	55·39	54·74	51·83
4—5	54·11	56·17	54·71	53·73
5—6	56·51	56·67	56·42	56·44
6—7	56·68	57·65	58·16	56·64
7—8	56·74	58·12	—	56·62
8—9	57·30	57·70	—	56·92
9—10	57·88	—	—	58·53

TABLE 14

Teat size (in mm)

Order of lactation		*Hariana*			*Tharparkar*		
		Teat			*Teat*		
		Length	*Diameter*	*Placement*	*Length*	*Diameter*	*Placement*
1.	**Front**	4·81	1·82	8·74	5·31	2·08	11·74
	Hind	4·30	1·52	5·00	4·95	1·85	6·21
2.	**Front**	5·06	1·87	9·09	5·50	2·26	11·99
	Hind	4·32	1·65	5·17	5·18	2·08	7·08
3.	**Front**	5·57	2·02	9·62	6·54	2·46	10·43
	Hind	4·74	1·82	5·85	6·09	2·23	6·58
4.	**Front**	5 66	1·94	9·09	6·46	2·58	10·23
	Hind	4·90	1·70	4·88	5 54	2·28	6·04
5.	**Front**	6·09	2·36	9·37	6·88	2·45	10·85
	Hind	5·58	2 15	5·37	6·11	2·40	6·71
6.	**Front**	5·66	2·09	10·50	7·43	2·53	8·93
	Hind	4·92	1·89	5·83	6·96	2·37	5·86
7.	**Front**	6·15	2·49	8·50	7·35	2·81	6·00
	Hind	5·53	2·24	6·00	7·15	2·62	3·75
8.	**Front**	5·60	2·73	—	7·90	2·73	9·16
	Hind	5·11	2.56	—	6·37	2·25	6·33
9.	**Front**	6·28	2·50	—	7·10	2 51	7·25
	Hind	5·81	2·30	—	6·40	2·12	6·25

Analysis showed that only in one out of the four *Hariana* and one out of two *Tharparkar* farms significant variation due to order of lactation was observed in case of fore teats. In case of hind teats, similar was the case in *Hariana* and in *Tharparkar* (both farms) exhibited significant variation due to order of lactation.

Similar results were seen in case of teat diameter and distance between front teats. There was no significant variation in case of distance between hind teats.

Order of lactation and teat measurements were found to be significant in some *Hariana* farms and not in others.

All the measurements in all the breeds showed steady increase and got stabilised after some age. In case of *Sahiwal* and *Sindhi*, it occurred at the age between 3-4 years, and in the *Hariana* and *Tharparkar*, above the age of 5 years. None of the measurements showed farm to farm variation within a given breed.

Estimates of hrritability obtained for the various measurements in the different farms and breeds are given in Table 11.

TABLE 11

Estimates of heritability of head size

Farm No.	*Head length.*	*Orbital length*	*Maxillary length*	*Length of ramus*
		Hariana		
1.	0·55±0·15	0·62±0·17	0·62±0·15	0·48±0·14
2.	0·56±0·16	0.54±0·16	0·60±0·17	0·48±0·14
3.	0·55±0·17	0.52±0·17	0·64±0·18	0·54±0·17
4.	0·48±0·13	0.52±0·14	0·59±0·15	0·63±0·16
5.	0·56±0·22	0.64±0·24	0·48±0·21	0·67±0·25
		Sahiwal		
1.	0·37±0·18	0·36±0·18	0·29±0·15	0·40±0 19
2.	0·78±0.27	0·46±0.20	0 16±0·14	0·40±0·19
3.	0·68±0·26	0·45±0·20	0·94±0·31	0·15±0·13
4.	0·72±0·35	0 95±0·42	0·23±0·22	0·95±0·42
		Sindhi		
1.	—0·09	1·14±0·54	1·33±0·61	—0·47
2.	0·33±0·18	0 61±0·25	0·68±0·26	0·48±0·22
		Tharparkar		
1.	0·42±0·12	0·51±0·13	0·40±0·12	0·54±0·14
2.	0·84±0·21	0·62±0·17	0·56±0·16	0·72±0·19

In the majority of cases, the estimates were high, showing the existence of large additive genetic variation for selection.

In case of *Sahiwal* and *Sindhi* cattle breeds, head length did not increase in proportion to orbital width, while in *Hariana* and *Tharparkar* orbital length increased or decreased proportionately with head length as observed from the behaviour of cephalic index with age.

Summary

A summary of the work done at the Animal Genetics Division of Indian Veterinary Research Institute, Izatnagar, under the guidance of the author on finding the extent and nature of variation present in certain physical traits in some of the more important cattle breeds of India is given. It is concluded, that the study of variation will provide a valuable guide for framing a rational and broad based policy of cattle improvement for this country.

Acknowledgment

I am thankful to my students for the use of the unpublished data contained in their Masters' theses in preparing the present article.

REFERENCES

Amble, V.N , Krishnan, K.S. and Srivastava, J.S. "Statistical studies on breeding data of Indian herds of Dairy cattle —I", *Indian J. Vet. Sci* , 1958a, **28**, 33-82.

Amble, V.N., Krishnan, K.S. and Soni, P.N. "Age at first calving and calving interval for some Indian herds of cattle", *Indian J. Vet. Sci.*, 1958b, **28**, 83.

Amble, V.N. and Krishnan, K.S. "Statistical studies on breeding data of Indian herds of Dairy cattle—II", *Indian J. Vet. Sci.*, 1960, 30, 1-29.

Amble, V.N., Krishnan, K.S. and Soni, P N. "A review of the breeding results obtained in some Indian herds of Dairy cattle", *Proc. 14th Meet. Anim. Husb. Wing. Bd. Agric. Anim. Husb. India* (Bangalore) 1963, 428-450.

Bhatia, S.S. "Studies on genetic parameters of economic traits in Zebu breeds of cattle II – Skin thickness in *Sahiwal* and *Red Sindhi*" M.V.Sc. thesis, Agra University, India, 1967.

Chatterjeo, B.C. "Studies on variation in head size of milch breeds of Zebu cattle", M.V.Sc. thesis, Agra University, India, 1967.

Chet Ram. "Skin thickness in *Hariana* cattle", Ass. IVRI thesis, Izatnagar, U.P., India, 1955.

Chet Ram. "Variability of skin thickness in *Hariana* cattle", *Indian J. Vet. Sci.* 1968, Sept. issue (accepted for publication).

Choudhury, R.R., "Studies on variation in head size of grey breeds of Zebu cattle", M.V.Sc, thesis, Agra University, India, 1957.

Hazel, L.N. and Terril, C.E. "Heritability of weaning weight and staple length in range *Rambouillet* lambs". *J. Anim. Sci.*, 1945, **4**, 347-358.

Indian Council of Agricultural Research, New Delhi, "Definitions of the characteristics of cattle and buffalo breeds in India", ICAR Bull. No. 86.

Joshi, N.R. and Phillips. R.W., "Zebu cattle of India and Pakistan", FAO Agri. Studies No. 19 (Rome), Italy, 1953.

Mohar, M. "Studies on genetic parameters of economic traits in Zebu breed of cattle-I —Skin thickness in *Hariana* and *Tharparkar*" M.V.Sc. thesis, Agra University, India, 1967.

Narayan, A.D., "Studies on variation in horn pattern in grey breeds of Indian cattle", M.V.Sc. thesis, Agra University, India, 1967.

Olver, A. "A survey of some of the important breeds of cattle in India". Misc. Bull. No. 17. Indian Council of Agricultural Research, New Delhi, India, 1938.

Randhawa, M.S., "Agriculture and Animal Husbandry in India", xiv+342, Indian Council of Agricultural Research, New Delhi, India, 1962.

Statistical analysis showed that while the udder lengths varied significantly in the two breeds, there was no significant variation in either farm or lactation within a breed.

In case of udder width, the difference between breeds was significant. That between lactation was significant in *Tharparkar* and one of the four *Hariana* herds. The difference between farms also showed significant variation.

Similar was the situation in respect of udder depth also.

6. Teat size and shape

Three types of teats were recognized. They were (*a*) cylindrical, (*b*) funnel and (*c*) bottle shaped. Their incidence in the 2 breeds are given in Table 13.

TABLE 13

Types of teats (%)

Breed		*Cylindrical* %	*Funnel* %	*Bottle* %
Hariana	Fore	76·2	22·8	1·0
	Hind	63·7	35.8	0·5
Tharparkar	Fore	77·0	14·9	8·1
	Hind	61·5	29·1	9·5

Cylindrical types were the most common, though considerable farm to farm variation was noticed which was significant within a breed. The incidence varied significantly in respect of the other two types between the two breeds. There was variation too in the incidence of the different types of teats in the fore and hind quarters.

Three measurements in respect of teat size were made. They were (*a*) teat length, (*b*) teat diameter and (*c*) placement of teats. The averges of these for the two breeds are given in Table 14.

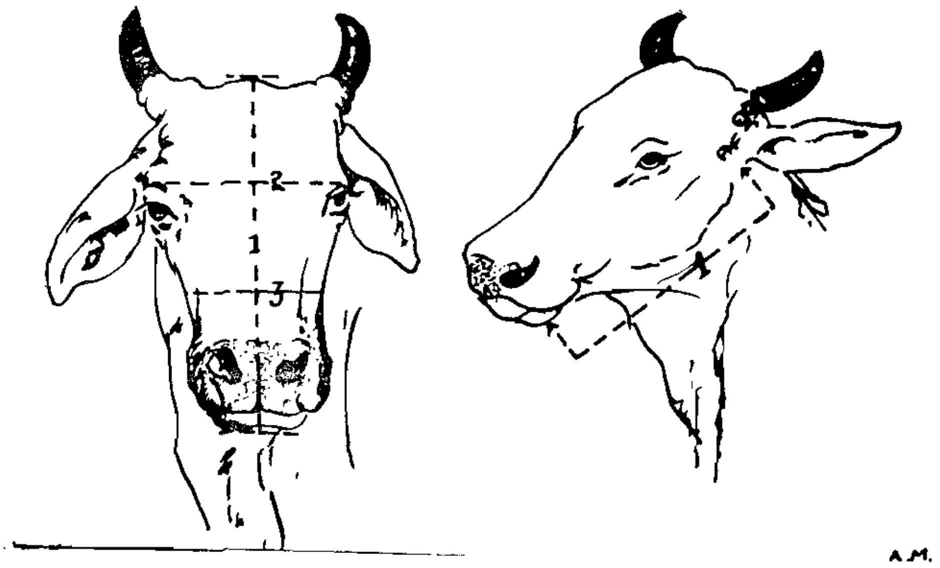


Fig. 1.

Front view
1. Length of head
2. Orbital width
3. Maxillary width

Side view
1. Length of ramus

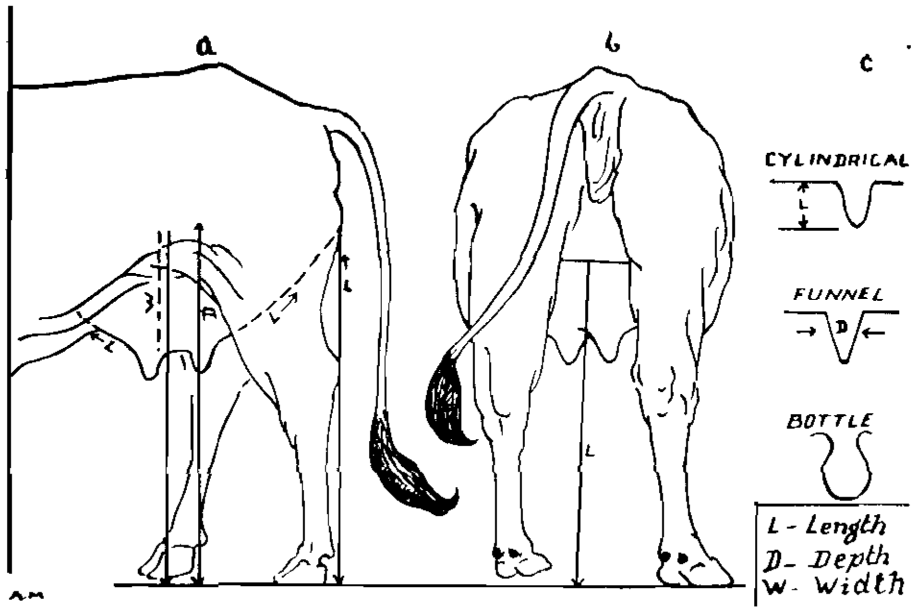


Fig. 2.

Discussion

The present studies have brought out the necessity of finding the extent of variation in physical traits actually present in our cattle breeds, and how much of that is available for exploitation through selective breeding. Only when we know what we have to select from, then only a rational plan of cattle improvement can be thought of and worked into the fabric of cattle improvement programmes. Unfortunately, for this country, the awaking after a thousand or more years of sleep or neglect in respect of cattle improvement, had come with a wrong bias. The initial emphasis had not been so much on finding the actual variation present, but rather on definition of breeds, definition of breed characteristics and opening of herd books on a priority basis, most on 'ad hoc' considerations, instead of basic fundamental ones. Without proper appreciation of the factors involved in the deterioration of the production capacity of our animals they had been condemned out of hand, and measures suggested for crossing them all with bulls of exotic breeds. While under proper conditions of nutrition, management and housing the crossbred cow is bound to behave better than the average Indian cow, there is no guarantee that they may continue to do so, when conditions change, and the self same animals are subjected to unfavourable conditions as for example in the villages of India of today. The indigenous cattle have already been subjected to these rigors of nutritional deficiency and climatic stress and the better of the Indian breeds, are producing whatever they do *in spite of the same.* A little help to them in the matter of better feed, management and disease control may give better returns, than embarking on a national policy of mongrelisation *irrespective of the fact whether the local conditions are conducive to the well-being and behaviour of the decidedly delicate and more susceptible crossbreds.* It is in this context that studies of the type now reported, and others in hand will provide useful pointers in shaping a more rational cattle breeding policy for this country.

Analysis carried out by Amble and his associates on farm data had shown some of the lacunae that hindered improvement on right lines. The most common being use of too many bulls, and failure to rear sufficient number of daughters from each one of them for effecting progeny test. The number of bulls used at a time too were too few so that chances of getting one or two of outstanding merit genetically were small. They had recommended the use of at least 10 at a time. Considering this and other factors, it is clear that whatever breeding on well run cattle farms that had taken place was not carried out on right lines. It would seem that the right solution lies in reorganisation of existing resources after knowing exactly what they are and removing the defects that hindered development earlier rather than independence on foreign countries for use of their breeds and experts.

THE USE OF CONSUMER'S AND PRODUCER'S SURPLUS IN THE EVALUATION OF PROJECTS APPLIED TO INDIAN AGRICULTURE*

By

GERHARD TINTNER[1]

AND

MALVIKA PATEL[2]

In modern welfare economics there has been a renewed interest in the Marshallian concept of consumer's and producer's surplus [1, 3, 10]. These simple concepts conceived in the great tradition of English utilitarianism give us a rough idea of the welfare aspects of planning, which, unfortunately,, are frequently neglected. National income or national income per capita gives an even rougher approximation. The target functions, used *e.g.* in planning in the Netherlands [5, 6, pp. 272-528, 7, 9] are very interesting but involve arbitrary simplifications. The use of shadow prices [4, pp. 43-54, 8] appears also somewhat questionable and certainly is very difficult to apply.

To give a justification of consumer's surplus, we consider a society which is composed of m individuals with similar tastes. The individual i has a utility function:

$$U_i = U_i\,(X_{i1},\ldots,X_{in}) \tag{1}$$

where X_{ik} is the amount of goods or service k available to individual i.

Let his money income be M_i and consider static conditions. Equilibrium will exist if U_i is a maximum. The necessary conditions for a maximum of utility are

$$\frac{\partial U_i}{\partial X_{ij}} = L_i p_j \qquad (j=1, 2,\ldots, n) \tag{2}$$

and

$$\sum_{k=1}^{n} p_k X_{ik} = M_i \tag{3}$$

In these formulae p_k is the price of commodity or service k and L_i is the marginal utility of money. We may solve the system for the individual demand functions of the various commodities :

---

*Research supported by the National Science Foundation, Washington D.C. We are obliged to Dr. V.G. Panse and Dr. G.R.Seth of the Institute of Agricultural Research Statistics, New Delhi, for supplying the data.

[1] University of Southern-California, Los Angeles, California.

[2] San Diego State College, San Diego, California.

Sharma B.D., "Studies on variation in shape and size of udder in grey breeds of Indian cattle". M.V.Sc. thesis, Agra University, India, 1967.

Singh, U.B., "Studies on supernumerary teats in Indian cattle", M.V.Sc. thesis, Agra University, India, 1965.

Singh, U.B., and Prabhu, S.S., "Supernumerary teats (SNT) in Indian cattle-I—Incidence", *J. Anim. Morpho. Physiol. (Baroda)* 1967a, **13**, 34-45.

Singh, U.B. and Prabhu, S.S., "Supernumerary teats (SNT) in Indian cattle-II—Inheritance", *J. Anim. Morpho. Physiol. (Baroda)* 1966b, **13**, 46-49.

Snedecor, G.W., "Statistical methods", Iowa State College Press, *Ames, Iowa, 1956.*

Sukhatme, P.V., "Statistical study of a breeding experiment with goats", *Indian J. Vet. Sci.*, 1944, **14**, 167-176.

Ware, F. "A further survey of some important breeds of cattle and buffaloes of India", Misc. Bull. No. 54, Indian Council of Agricultural Research, New Delhi, India, 1942,

The sum of consumer's and producer's surpluses is given by

$$(14) \qquad R(s)=\int_0^{q_1} f(q,\, s)\; dq+\int_{q_1}^{g_0}\Big[\; f\,(q,\, s)-g(q,\, s)\Big]dq.$$

If the shift parameter is continuous, we have for the derivative

$$(15) \qquad R'(s)=f(q_1,\, s)(\partial q_1/\partial s)+\int_0^{q_1}\Big[\; f(q,\, s)/s\;\Big]dq.$$

$$+\Big[f(q_0,\, s)-g(q_0, s)\Big](\partial\, q_0/\partial s)-\Big[f(q_1,\, s)-g(q_1,\, s)\Big]\frac{\partial q_1}{\partial s}$$

$$+\int_{q_1}^{q_0}\Big[\;\partial f(q,\;\, s)/\partial s-\partial g(q, s)/\partial\, s\Big]dq.$$

Now we might identify the shift parameter s with a project. For example, a project might lower the supply function, $p=g(q,\, s)$, because of such changes as the introduction of new production methods or the training of labor. But a project could also shift the demand function $P=f(q,\, s)$ upwards, because of such factors as propaganda or the additional income created.

Given an initial situation with $s=S_0$, we might compare the difference

$$(16) \qquad R(S_1)-R(S_0)$$

with the total cost of the shift from S_0 to S_1. Also, we might compare the discounted value of (16) with the total discounted costs of bringing about the shift.

If the project is large, it will involve, of course, changes in demand and supply functions of other commodities. In principle, we should try to estimate the effect of shifts on other markets than the one concerned, but this seems to be a very difficult task. In a summary fashion, this might be accomplished by considering the effect of the shift on the demand function, by means of a multiplier effect.

This procedure has been justified by Winch (11, p. 421) as follows :

"If the sum of consumers' gains, measured by the changes in areas under the relevant market demand curves, exceeds the sum of cousumers' losses similarly measured, the policy should be adopted unless the redistribution involved is held to be sufficiently unfavourable to nullify this conclusion.

If the sum of consumers' losses exceeds the sum of the gains, the policy should not be adopted unless the redistribution involved would be sufficiently favourable to outweigh the gain criterion.

(23) $$p = 7{\cdot}1345 - 98{\cdot}5475q$$

where q denotes per capita quantity consumed (in tons) and p denotes harvest price per ton deflated by cost of living index.

(24) $$q = {\cdot}01852 + {\cdot}001590I + {\cdot}001405\mathrm{F}$$

where I (proportion of irrigated area under wheat to total area under wheat) and F(amount of nitrogenous fertilizer used per acre for wheat) are shift parameters.

The equilibrium quantity is

(25) $$q_0 = {\cdot}01852 + {\cdot}001590I + {\cdot}001405\mathrm{F}$$

Equilibrium price is

(26) $$p_0 = 5{\cdot}3094 - {\cdot}1567I - {\cdot}1385\mathrm{F}$$

and the producers' and consumers' surplus is given by

(27) $$R(s) = \int_0^{{\cdot}01852 + {\cdot}001510I + {\cdot}001405\mathrm{F}} (7{\cdot}1345 - 98{\cdot}5475q)dq$$

$$= {\cdot}1152 + {\cdot}008434I + {\cdot}008738\mathrm{F} - {\cdot}0002202IF - {\cdot}0001246I^2 - {\cdot}00009727F^2$$

From these surplus functions (22) and (27) we can compute the surpluses with different amount of I and F. The following Tables I and II show the estimated surplus in rice and wheat respectively with different hypothesis about I and F. K_I and K_F refers to the levels of I and F in the year 1962-63 respectively.

TABLE I

Consumer's and Producer's Surplus for rice

A :	Both I and F increased	
I	F	R(S)
K_I = ·3803	K_F = 6·53	·4883
$1{\cdot}5K_I$ = ·5705	$1{\cdot}5K_F$ = 9·78	·71962
$2K_I$ = ·7606	$2K_F$ = 13·06	·94461
$3K_I$ = 1·1409	$3K_F$ = 19·54	1·37747
$4K_I$ = 1·5212	$4K_F$ = 26·12	1·78736
$5K_I$ = 1·9015	$5K_F$ = 32·65	2·17416
$10K_I$ = 3·8030	$10K_F$ = 65·30	3·76388

$$(4) \qquad X_{ik}=f_{ik}(p_1,\dots,p_n, M_i),\ k=1, 2, \dots, n.$$

The total demand functions are given by

$$(5) \qquad X_k= \sum_{i=1}^{m} f_{ik}(p_1,\dots,p_n, M_i).$$

Let us now return to equation (2). If we assume all other prices and incomes constant, then we have evidently

$$(6) \qquad U_i=L_i \int_0^{x_{ij}} p_j\, dx_j,$$

where we may now think of the expression p_j as a demand function, *i.e.* as the inverse of (4). Also, we assume L_i(marginal utility of money) constant and the same for all individuals. This assumption excludes the income effects.

By analogy, we might solve the collective-demand functions (5) in terms of the prices. The analogous expression

$$(7) \qquad U= \int_0^{x} j\, p_j\, dx_j,$$

might under similar assumptions be considered an approximation to total utility. This procedure assumes that all the individuals in the community are similar, at least so far as marginal utility of money is concerned.

Now consider a market for a given commodity. Denote the price of the commodity as p and its quantity as q, and let s be a shift parameter.

Then

$$(8) \qquad p=f(q, s)$$

is the demand function. For simplicity, we assume it continuously falling. The supply function is

$$(9) \qquad p=g(q, s).$$

It is assumed not decreasing. Define by

$$(10) \qquad g(q, s)=0$$

a function

$$(11) \qquad q_1=q_1(s)$$

which shows where the supply function crosses the q exists. Also we have

$$(12) \qquad f(q, s)=g(q, s)$$

which defines the equilibrium quantity

$$(13) \qquad q_0=q_0(s).$$

B. Only I increased, F constant at 1962-63 level

I	F	$R(S)$
K_I	K_F	·15577
1·5K_I	K_F	·15698
2 K_I	K_F	·15819
3 K_I	K_F	·16061
4 K_I	K_F	·16296
5 K_I	K_F	·16529
10 K_I	K_F	·17653

C. Only F increased, I constant at 1962-63 level

I	F	$R(S)$
K_I	K_F	·15577
K_I	1·5K_F	·17238
K_I	2 K_F	·18757
K_I	3 K_F	·21347
K_I	4 K_F	·23347
R_I	5 K_F	·24761
K_I	10 K_F	·23001

The results show that greater use of fertilizer and irrigation can increase the surplus.

If the redistribution criterion does nullify the decision criterion indicated by the gain criterion, consideration should be given to a change in distribution policy which would make any remaining redistribution acceptable."

We attempt to apply the analysis to Indian 'Agriculture, considering use of fertilizer and irrigation as shift parameters (1950—63). The demand function for per capita quantity (in tons) consumed of rice is estimated as

$$(17) \qquad p = 6{,}5455 - 4\cdot0878q$$

where p is the real price *i.e.* harvest price per ton deflated by the cost of living index.

Regarding the supply function no systematic relation was observed between price and quantity supplied. Hence we assume supply as completely price inelastic. The estimated supply function for rice is

$$(18) \qquad q = \cdot001535 + \cdot1895\mathrm{I} + \cdot0004453F$$

where I and F (proportion of irrigated area under rice to total area under rice, and amount of nitrogenous fertilizer used per acre for rice) are shift parameters.

We derive the equilibrium quantity

$$(19) \qquad q_0 = \cdot001535 + \cdot1895I + \cdot0004453\mathrm{F}$$

which is the same as the quantity supplied—supply being perfectly price inelastic.

The equilibrium price is derived by substituting the equilibrium quantity in the demand function,

$$(20) \qquad p_0 = 6\cdot5392 - \cdot7665I - \cdot00120\mathrm{F}$$

The sum of the consumer's and producer's surplus is given by

$$(21) \qquad R(s) = \int_0^{q_0} f(q)dq$$

Here we have

$$(22) \qquad R(s) = \int_0^{\cdot001535+\cdot1895I+\cdot0004453F} (6\cdot5455 - 4\cdot0878q)\, dq$$

$$= \cdot01005 + 1\cdot2392I + \cdot002912\mathrm{F} - \cdot0003450I\mathrm{F}$$

$$- \cdot07340I^2 - \cdot0000004053\mathrm{F}^2$$

Similarly we have demand and supply functions for wheat estimated as (23) and (24) respectively.

DEVELOPMENT OF SOIL STUDIES IN INDIA FOR INCREASED AGRICULTURAL PRODUCTION WITH PARTICULAR REFERENCE TO SOIL SURVEY, SOIL TESTING AND FERTILIZER TRIALS

by

S.P. RAYCHAUDHURI*

Low Fertility of Indian Soils

The Royal Commission on Agriculture in its Report expressed the view that the fertility of Indian soils is as low as it was at the time of death of Akbar (1602). The stability of fertility of Indian soils at a low level is apparently due to the natural processes of recuperation of nitrogen from the atmosphere by the soil microflora. There had, however, been indication from the Agricultural Statistics, until last five to six years that the yield per hectare of rice and wheat has tended to decrease, somewhat. Since last five to six years, however, the yield per hectare in respect of these two important food crops has shown tendency to increase. This has been possible to a very large degree, due to better soil management practices. However, it is essential that the fundamental and applied work on the maintenance and improvement of soil fertility in its different aspects should not only be kept up but intensified more and more. Since last quarter of a century considerable work has been carried out on Indian soils on the improvement of soil fertility and on the treatment and reclamation of problem soils.

Plant Nutrients

It has been estimated that cultivated crops remove annually from Indian soils about 4·2 million tonnes nitrogen, 2·1 million tonnes of phosphorus as P_2O_5 and 7·3 million tonnes of potash as K_2O and 4·8 million tonnes of lime as CaO. In addition, there is depletion of plant nutrients due to soil erosion. With limited quantities of artificial fertilisers at our disposal, utilisation of all indigenous materials like farmyard manure, compost, bone-meal, basic slag, non-edible oil cakes, green manures, etc. are urgently necessary. Also for assessment of nutrient status, detailed knowledge of the soils of the country is essential. Intensive work on these lines is being carried out in the States, the All India Soil and Land Use Survey Scheme and by the Indian Council of Agricultural Research through its various Central Research Institutes in Agriculture.

---

*Planning Commission,

REFERENCES

1. Hicks, J.R. A Review of *Demand Theory*, Oxford, Clarendon Press, 1956.
2. Little, I.M.D. A Critique of Welfare Economics, Oxford, Clarendon Press, 1950.
3. Mishan, E.J. "A Survey of Welfare Economics, 1939-59", Econ. J. 70 : 197-256, June, 1960.
4. Radner, R.R. Notes on the Theory of Economic Planning Atehns, Serihnus Press, Center of Economic Research, 1963.
5. Theil, Henri. "Econometric Models and Welfare Maximisation," Weltwirtschaftliches Archiv. 72 : 60-81, Hft, 1, 1954.
6. Theil, Henri. Economic Forecasts and Policy, Amsterdam, North-Holland Publishing Company, 1961.
7. Timbergen, Jan. On the Theory of Economic Policy, Amsterdam, North Holland Publishing Company 1952.
8. Timbergen, Jan. The Design of Development, Baltimore, The John Hopkins Press, 1958.
9. Timbergen, Jan. Economic Policy, Principles and Design Amsterdam, North-Holland Publishing Company, 1956.
10. Tintner, G. and Patel, M. "Evaluation of Indian Fertilizer Projects: An application of Consumers' and Producers' Surplus". Journal of Farm Economics, Vol. **48**, 704-710, August 1966.

    "Correction" ibid. Vol 49. May 1967.
11. Winch, David M. "Consumer's surplus and the Compensation Principles." Am. Econ. Rev. 55 : 395-423, June 1965.

B. Only I increased ; F constant at 1962-63 level

I	*F*	*R*(*S*)
K_I	K_F	·41883
1·5K_I	K_F	·71080
2K_I	K_F	·92734
3K_I	K_F	1·34469
4K_I	K_F	1·74081
5K_I	K_F	2·11564
6K_I	K_F	3·67156

C. Only F increased ; I constant at 1962-63 level

I	*F*	*R*(*S*)
K_I	K_F	·48883
K_I	1·5K_F	·49787
K_I	2K_F	·50695
K_I	3K_F	·52504
K_I	4K_F	·54309
K_I	5K_F	·56104
K_I	10K_F	·65054

TABLE II

Consumer's and producer's surplus for wheat

A. Both I and F increased

I	*F*	*R*(*S*)
K_I = ·3413	K_F = 5·50	·15577
1·5K_I = ·5120	1·5K_F = 8·25	·17353
2 K_I = ·6826	2 K_F = 11·00	·18957
3 K_I = 1·0239	3 K_F = 16·50	·21663
4 K_I = 1·3652	4 K_F = 22·00	·23694
5 K_I = 1·7065	5 K_F = 27·50	·25052
10 K_I = 3·4130	10 K_F = 55·00	·21732

Ground Water

An important line of work on which attention has been focussed in recent years in the country is the quality of ground water and irrigation water and their effect on the salinity and alkalinity of the soil. Occurrences of salinity leads to deterioration of soil structure and soil fertility. Such problems are acute, particularly in the Punjab, Uttar Pradesh, Rajasthan, Gujarat and Maharashtra. Methods of reclamation of saline, alkaline and water-logged soils and knowledge of the tree, shrub and grass species occurring in such areas are generally known. But areas under sweet and saline ground water have not been demarcated, except in very limited cases. Intensive work on this problem is being carried out by the Geological Survey of India, the Exploratory Tube Well Organisation, the State Governments and the Indian Council of Agricultural Research.

Crop Rotation

The usefulness of the inclusion of a legume in the rotation has been established. However, crop rotation including legume has been favoured by the farmers to a limited extent only. In the Punjab berseem (Egyptian clover) and some other legumes have found favour with some peasants. The question of systematic studies of legumes and their introduction is at present receiving attention. There is immense scope for the development of the use of legumes and grasses and other types of fodder as a measure of improving soil fertility and controlling erosion. The fact remains, that the use of legumes being very limited, the effect on the general level of productivity of soil has not yet been marked. Investigations with indigenous and exotic leguminous plants and grasses with a view to testing their suitability for different purposes, it is hoped, will soon be intensively pursued.

Micro-nutrients

The essentiality of a group of six elements *viz. B., Co., Cu, Mn, Mo* and *Zn* in the nutrition and metabolic activity of plants and animals has been realised by Indian workers in comparatively recent years. Under certain conditions of soil and climate, the amounts of these elements absorbed by plants from soil are not enough to supply normal requirements. Alkaline, overlimed, calcareous, strongly acidic, highly leached, eroded, sandy and peaty soils are likely to be deficient in the above elements and such soil conditions are not uncommon in India. The widespread nature of deficiencies of these elements in our crops and soils has also been recognised. No systematic work on the distribution of these elements in Indian soils has however been carried out and the nutritional requirements of these elements by different crops have not also been properly assessed. Investigations on the deficiencies of these micronutrients in relation to different soil-climate crop conditions require intensification and the Indian Council of Agricultural Research has been financing coordinated scheme on micronutrients through which very useful information has already been obtained.

Clay Minerology

The clay is the most reactive portion of the soil and by virtue of its powers of adsorption and ionic exchange capacities it controls the chemical and physical properties of soil, besides holding all the important elements of plant nutrition. Kaolinitic clays are known to fix large amounts of soluble phosphate from added fertilisers, whilst montmorillnitic clays, under conditions of alternate wetting and drying, and illitic clays, under continuously moist conditions, fix large amounts of potassium and ammonia from the fertilisers. The free oxide also plays an important role in phosphate fixation as does vermiculite in the case of K and NH_4 fixation. The physical and chemical properties of soils ultimately depend upon the nature and amount of clay minerals present. The work on clay minerals is a very specialised type of work and is being carried out at the Indian Agricultural Research Institute and some State agriculture laboratories and the Universities. In future investigations, the mineralogical make up of the sands and clay fractions should play a prominent part in the classification of Indian soils.

Soil Microbiology

In spite of the attention of numerous agricultural scientists of the country, the causes and extent of nitrogen recuperation in soils of different parts of the country remain still unknown. The possibility of improvement of pulse yields by cross-inoculation with various organisms from wild legumes still remains. Much work still needs to be done on the completion of effective and ineffective strains of rhizobium of common cultivated legumes in our soils. With the increased use of fertilisers in the country, it is essential that knowledge be obtained regarding their effect on the microbial population of the soil. Maintenance of a healthy microbial population in the soil is very necessary for continued fertility of our soils. It may be necessary for soil microbiologists to devote some time and work on gaining information on the above aspects so that definite systems of fertiliser practices can be developed for different parts of our country.

There is an inverse relationship between mean annual temperature and the level of organic matter in the soil in regions of comparable rainfall. Higher temperatures stimulate microbial decomposition more than they stimulate the production of plant tissue. This relationship has great practical importance as a guide in establishing feasible limits at which levels of organic matter can be maintained. Ten to twenty kilograms of plant residues are required to provide energy for producing one kilogram of soil organic matter. The treatment of the soil with plant residues enhances the rate of humus breakdown. This is due to the large microbial population that builds up, in response to ådded substrate and brings about greater rate of decay. This explains some of the field observations of lack of organic carbon accumulation following the ploughing of green manure crops. In grass lands the excretions of extensive root network may provide an adequate supply of available carbohydrates and under such conditions non-symbiotic nitrogen fixation is of practical consequence. Further, the rhizosphere of crops is a site of intense metabolic activity and organic matter is continuously formed and mineralised and appreciable non-symbiotic fixation of nitrogen may take place. There is a great need to re-examine the situation, by seed inoculation with active strains of non-symbotic nitrogen fixing microflora and providing favourable conditions for maximum fixation. Moreover, some of these organisms may benefit the crop by way of synthesis of growth promoting substances as is the case with Azotobacter. Work on some of the above important aspects is underway at the Indian Agricultural Research Institute and State Agricultural laboratories.

In rice fields and flooded areas, nitrogen fixation by blue green algae is appreciable. In contrast to the heterotrophic bacteria, the blue green algae are not restricted in their carbon nutrition to performed organic matter, since their carbon source is carbon dioxide. These may add nitrogen in denuded soils and barren areas. There is need to carry out intensivestudy of the blue green algae and select suitable efficient strains and examine the possibility of increasing crop yields through algal inoculation.

Nitrogen fixation through legumes is of considerable importance varying between 44·84 to 224·20 kg per hectare. Detailed studies need be conducted on the conditions of maximum benefit of legume seed inoculation with the relevant rhizobium strains and on the preparation of mass cultures of the same. Critical work also needs to be conducted on the isolation of active strains of phosphorus solubilising organisms and study of the conditions, for maximum benefit of seed inoculation with these, if any. Some work on the above important problems is underway at the Indian Agricultural Research Institute and some State Agricultural laboratories.

Soil Organic Matter

The direct and indirect effects of soil organic matter in the maintenance of soil fertility have been recognised. Investigations on some fundamental aspects of soil organic matter are underway. In some soils fertilized with organic manures 0·22 mg per kg of soil of auxin has been observed while in unfertilised soil the value is 0·66 mg/kg soil and in soils with mineral fertilisers, the value is 0·09 to 0·106 mg/kg soil. It is necessary to examine under Indian conditions the role of organic matter in the synthesis and accumulation of vitamins in the soil and to study the type of microflora that takes part in it.

Radio techniques in soil fertility investigations

In recent years, tracer techniques is finding much use in understanding the various metabolic reactions in plants, the movement, fixation and changes in plant nutrients in the soil, the exact evaluation of the available nutrients In soil and in the determination of fractions of the added fertilisers actually utilised by the plants. In many cases, effectiveness of the fertiliser depends on how it is placed in the soil in relation to the crop lands. Using P 32 labelled single superphosphate it has been shown at the Indian Agricultural Research Institute that crop species differ in the utilisation of added phosphorus but also that it differs from soil to soil for the same crop species. It has also been shown that 50-60 per cent of our soils are deficient in phosphorus. A method of estimation of available phosphate in the soil has also been developed. It was also found that unsaturation of exchange capacity would be a good index of utilisation of applied phosphorus.

Soil Structure

The capacity of any soil for the growth of plants and its response to management depends as much on its fertility as on its structure. The importance of soil structure in soil classification and in influencing soil productivity can scarcely be over-emphasised. Soils with aggregates of spheroidal shape have much more space between aggregates. have more rapid permeability and are more productive than soils of comparable fertility that are massive or even coarsely blocky or prismatic. The poor yielding quality of the black soil (regur) of Central India has been traced to a loss in soil structure resulting from bad drainage and water-logging. Recent development of certain soil conditioners has brought to light the factors associated with the agregation of soil particles. Regur with high clay content has been found to produce bigger sized water-stable aggregates under treatment of soil conditioners like krilium and tamarind seed powder.

Soil Moisture

The relation between water and a complex body like the soil is one of the most complicated problems in soil physics. Soils show great variation in their power to hold moisture, the depth to which moisture may percolate and the degree to which desication can take place. All these factors should be properly understood for Indian soils in order to have an efficient agriculture. The complications multiply in the field where conditions are daily changing and where meteorological conditions over which no control is possible influence the movement of soil moisture to a very great extent. Some work in this line is underway but intensification of the work is necessary.

Soil Conservation

The modern concept of soil conservation embraces not only erosion control but also adoption of all round soil management and soil fertility improvement practices that will contribute to achieve sustained high yields from every hectare of agricultural land. Every year millions of tonnes of valuable surface soil is washed away during the monsoon. It has been estimated that about 80 million hectares of lands in the country requires protection by soil conservation measures. Amongst the various soil conservation measures evolved at the soil conservation Research Stations in the country particular mention may be made of the control of gullies by check dams and control of soil erosion on the hills by construction of terraces. As a typical example of increase in soil fertility achieved through soil conservation measures mention may be made of the spectacular increase in production at the Rehmankhera farm in Uttar Pradesh. The need for soil conservation practices, however, is not adequately realised by the cultivators and follow up action is necessary to preserve the achieved results.

Soil Classification and Soil Survey

Soil classification gained prominence in India when lands were assessed for revenue in the sixteenth century. In this system of land classification, stress was laid on certain inherent soil characteristics and external features, viz. texture and colour of soils, slope of land and availability of water, but the ultimate criterion was the yield of crops. On the basis of information thus collected and of considerations of marketing facilities, fair estimates of land values were arrived at.

In the beginning of the present century, attention was focussed on improving the productivity of the land and classification of land based on the contents of plant nutrients *viz.* nitrogen, phosphorus and potassium and also of lime of the top fifteen centimeters of soil was undertaken in selected areas. Land classification on the basis of mechanical composition of surface soils was also carried out in a few cases. Such classifications were of the nature of soil reconnaissance and had considerable agronomic importance. This method looked upon the surface soils as the primary source of plant nutrients and have been extensively used in most soil survey work, mainly during pre-1928 period and also in a few cases later on. In these studies no standard and uniform methods of sampling and analysis were used. Therefore, the conclusions obtained have limited value. Moreover, in this system of classification there is a great lack of coordination and understanding of the genetic relationship of various soils, so that the results of one experimental farm cannot be applied to another with a degree of certainty.

The Irrigation soils surveys which are carried out are either pre-irrigation or post-irrigation surveys. The purposes of the pre-irrigation soil survey are (*i*) allocation of areas fit for cultivation as soon as water becomes available, (*ii*) identification of lands which can be reclaimed and determination of the amount of water required for reclamation, (*iii*) determination of the prospects of rise in water table and water logging and (*iv*) provision of drainage. The object of post-irrigation soil survey on the other hand, are (*i*) to locate the areas that have already deteriorated or are in the process of deterioration, (*ii*) to formulate methods for the reclamation of deteriorated areas, and the saving of those which are likely to deteriorate, and (*iii*) to determine the type of drainage required for reclamation or prevention of deterioration. For such surveys, the surface and sub-soil samples are analysed for their mechanical constituents, *i.e.* clay and silt, soluble salts and exchangeable bases. The soils are often examined to a depth of about 90 cms. or more but the morphological characteristics of the soil profiles are not always taken into consideration. Amongst the important pre-irrigation soil surveys which are in progress the following may be mentioned : (*i*) Bhakra-Nangal Project ; (*ii*) Damodar Valley Project ; (*iii*) Nagarjunasagar Project ; (*iv*) Rajasthan Canal Project ; (*v*) Hirakud Dam Project ; (*vi*) Mahanadi Delta Irrigation Project ; (*vii*) Chambal Project ; (*viii*) Tungabhadra Project; (*ix*) Ghod Project ; (*x*) Dantiwada Project ; (*xi*) Mayurakshi Project ; (*xii*) Kangsabati Projec ; (*xiii*) Ramganga Project ; (*xiv*) Lower Bhawani Project and (*xv*) Kundha Project.

Soil survey for agronomic purposes was started by Basu and Sirur at the Padegaon Sugar-cane Research Station in the Bombay Deccan and the results were published in 1938 (Basu and Sirur, 1938). The black soils of the Decean were classified into twelve types on the basis of the genetic system of soil classification originally developed by the Russian pedologists, and adopted sucessfully by the American and European soil scientists. The work of Mukherjee and Dass (1943) and of Agarwal and Mehrotra (1952, 1958) in Uttar Pradesh showed that a consideration of the genetic factors in soil survey work can lead to a better understanding of the soils in regard to their characteristics as well as management for optimum production. Similarly, the genesis of red and laterite soils was studied by Raychaudhuri *(1941), Satyanarayana and Thomas (1961), and Biswas and Gowande (1962).*

Survey for soil conservation and land use has been undertaken in the country mainly during the post-independence period with the object of putting the land for optimum use. The different purposes for which such surveys are carried out are as follows :

(*i*) Adoption of soil conservation measures in the catchment areas of river valley projects for the purposes of, or from the point of view, of decreasing siltation in the reservoirs and increasing the life of the dams.

(*ii*) Survey of wastelands from the point of view of classifying the lands into those which can be put under cultivation economically and those which cannot be under cultivation economically but should be put to other uses such as pasture or afforestation.

(*iii*) Rehabilitation from the point of view of providing land for residential and cultivation purposes to families who have been uprooted from their homes either due to partition of India or due to their houses and lands having been immersed under water in the reservoirs made by the irrigation projects.

(*iv*) Soil survey for other soil conservation purposes like gully control.

Soil Survey has also been useful for finding out the suitability of soils for construction of highways, runways in the aerodromes, dams in the river valley projects, etc. Information regarding certain physical properties of the soil, such as cohesion, liquid limit, plasticity index, etc. is necessary for making highways, rail roads, air ports, towns and for making dams on rivers etc. These data are plotted on soil maps which contain the basic soil survey data.

Standard Soil Survey : Their interpretation

Standard Soil Survey, with varying scales of mapping and details of classification serve all purposes of soil use. The results of standard soil survey can be utilised in soil conservation, irrigation, agricultural extension, afforestation with minor additional information wherever necessary. Such standard soil survey take into account all soil characteristics that influence the various uses to which the soil can be put.

Standard detailed soil surveys are being carried out in the country on village maps of 1 : 3969 (16″ = 1 mile) or on any convenient scale, depending upon the availability of the maps. These field maps are later reduced to a scale 1 : 31680 (2″ = 1 mile) or 1 : 15840 (4″ = 1 mile) for publication and appending with soil survey reports which can be put to ready use by extension staff, soil conservationists or conservation engineers.

Soil character, depth, slope, land use or cover and degree of land susceptibility to erosion are the basic physical factors involved in sound planning for land use. These basic data are provided by standard detailed survey even for small areas and all physical variations that definitely affect erosion potentials and adaptable types of land use are indicated on maps.

The soil survey reports have found considerable use particularly in planning the distribution of water for irrigation farming, the drawing up of improved cropping patterns and in the preparation of plans for soil conservation. Thus the pre-irrigation reports of soil survey conducted in the Madhya Pradesh district of Mysore State have been made full use of in drawing up Block pattern of a triennial crop rotation ; the reports of the soil survey of the Tungabhadra Project and of the Nagarjunasagar Project have been used by the irrigation department for demarcating the zones which would be put under wet cultivation, and also zones which would come under a dry-cum-wet cropping patterns ; the soil survey reports published by the All India Soil and Land Use Survey Organisation have been used by State agencies for drawing up programmes of detailed soil conservation practices.

During the last decade a number of States started soil survey as an integral part of their agricultural development programme. The All India Soil Survey Scheme was started in the year 1956 at the Indian Agricultural Research Institute with soil correlation centres in four major soil regions at Delhi, Nagpur, Bangalore and Calcutta for carrying out reconnaissance soil survey throughout the country and for soil correlation. This scheme was integrated with the scheme of Land Use Planning Organisation of the Central Soil Conservation Board in the year 1958. Under the integrated scheme standard soil survey of critical areas as determined by aerial photo interpretation (on 1 : 3960, 1 : 7929 cadastral maps) is being carried out in the catchment areas of the major river valley project viz. Hirakud, Machkund, Chambal, Bhakra-Nangal, Ghod, Mayurakshi, Kangasbati, Dentiwada and Tungabhadra Projects as the requirements of soil information for soil conservation purposes in these catchment areas have been very great for decreasing the rate of silting of the dams. Standard soil survey is also being carried out in areas other than the catchments of the above dams where land development for agriculture, pastures, forestry. reclamation, rehabilitation has been projected by State Governments on priority basis. Reconnaissance soil survey is also carried out throughout the country for soil correlation purposes. Since the inception of the scheme till March 1967, a total area of about 10·6 million hectares have been covered by the All India Soil and Land Use Survey Organisation by detailed and reconnaissance surveys in the catchment areas of the aforementioned dams and in other areas in the States of Madras, Mysore, Kerala, Andhra Pradesh, Maharashtra, Orissa, West Bengal, Tripura, Assam, Uttar Pradesh and Punjab. A number of States who took up soil survey as a programme of soil and agriculture development covered till March 1967 about 20·0 million hectares partly by detailed survey but mostly by reconnaissance survey. In addition, the Central Water and Power Commission has carried out soil survey in the areas under the command of irrigation projects for about 11·5 million hectares till end of September, 1967. The total geographical area of the country is 326 million hectares. Therefore only about 13 per cent of the area of the country has been covered by soil survey which however is mostly of reconnaissance nature.

For the purpose of uniformity of soil and land use survey throughout the country a Soil Survey Manual was brought out in 1960 for serving as a guide to the techiques and procedural aspects of soil survey and soil mapping, which has since been revised.

The preparation of soil fertility maps and soil test summaries have been of help to planners, fertiliser industry and fertiliser production, distribution and consumption in the country. The All India Soil Test summaries prepared in the basis of 246134 samples analysed by all centres in 1965 showed that 54·3% of soil were low, 30·3% medium and 15·4% high with respect to available phosphorus ; 43·4% of soil were low, 33·7% medium and 22·9% high with respect to available potassium ; and 57% of soils were low, 33% medium and 10% high with respect of nitrogen.

Under Indo-U.S. Project, on soil fertility and fertiliser use, 24 soil testing laboratories have been set up at different locations. Besides the above, the Government of Bihar has established three new laboratories of their own at Ranchi, Patna and Pusa. Similarly State soil testing laboratories have been set at Palampur and Hissar. Till 1944, 43 soil testing laboratories are working at these different centres. These laboratories have been equipped so as to analyse generally 10,000 soil samples a year. More recently the laboratories at Ludhiana, New Delhi, Bangalore and Sambalpur have been remoddeled with the help of U.S. aid and every one of these laboratories is capable of analysing 30,000 samples a year. These laboratories have been rendering free advisory service to farmers. Till the end of the year 1966 about 1·38 million soil samples have been analysed in the soil testing laboratories. Under the All India Coordinated Scheme of the Indian Council of Agricultural Research some well equipped centres in the various soil climatic regions have been strengthened for working on the soil test and crops correlation work more intensively with the coordinating centre at IARI to evolve a firm basis for this purpose.

Although only a few years old, this programme has shown considerable progress. Soil fertility maps prepared on the basis of soil test data, have shown that use of phosphatic fertilisers is needed more in large areas where a major portion of soil samples (more than 50%) tested low in available phosphate. Need for nitrogenous fertilisers is almost universal. Similarly, problem soils testing acidic or alkaline require lime practices or application of gypsum respectively. Not much attention was given earlier to the problem of liming the soils in this country, and the need for taking suitable measures for these problem soils, is very urgent for stepping up crop production.

Correlation of soil survey, soil tests and fertiliser trials to crop responses

While soil survey determines the inherent characteristics of soils from which the reserve plant nutrients may be predicted, based on parent materials and other conditions of soil formation, the available nutrient contents on which crop growth mainly depends may vary even in the same soil type from place to place which is determined by soil testing. Soil survey and soil testing are thus closely inter-related and the results of soil tests along with the information from soil survey are useful for obtaining maximum return from the land.

Simple fertiliser trials in cultivators' fields throughout the country and the complex experiments at selected centres under different soil-climatic regions as also extensive soil tests carried out in the different regions have established the fertiliser needs of crops in different soil climatic regions. Further experiments in cultivators' fields are under-way for arriving at the most efficient fertiliser formulae in relation to soil and climate for the major crops.

Field experiments from the basis for formulating recommendations to the farmers on important agricultural techniques such as use of fertilisers and manures varieties etc. In the past, such experimental works have been mainly confined to agricultural experimental stations. The number of experimental stations in a country is usually small and further, the fertility of the soil and level of management at experimental stations are generally superior to those in cultivators' fields. Any generalisation of the conclusions obtained at experimental farms for applications over the country is, therefore, attended with risk. A concrete illustration of this risk is provided by the lack of response to phosphate indicated at experimental stations in India, while in large areas in farmers' fields, response to this fertiliser has been observed in recent trials. Such experimentation has also got considerable demonstration value and is, therefore, an essential step in persuading the farmers in other underdeveloped countries to adopt the use of fertilisers and other improved practices. An experimental programme with this objective must be spread over an adequate sample of farmers' fields and carried out with their cooperation and assistance. A representative sample of fields, selected randomly, will include a major proportion of fields belonging to farmers with limited means and small holdings. For the success of such a programme, the experiments to be carried out must satisfy the following conditions (Panse and Abraham, 1960).

(*i*) The experiments must be simple and consist of a small numher of plots.

(*ii*) The treatments which are to be superimposed on the farmer's normal practice, should be capable of providing critical information on main effects and interaction of different factors. The farmer's normal practice should also be included as one of the plots to serve as control.

(*iii*) The set of treatment in each experiment should have a demonstrating value. Each such set should therefore be self-contained in the sense that easily intelligible comparisons of practical value can be made from the experiment. All comparisons should be available with the farmer's normal practice as control.

(*iv*) The treatments should be promising as judged from past research at experiment stations, so that the possibility of the farmer incurring a loss by cooperating in carrying out the experiment should be minimum.

A fairly extensive programme of fertiliser trials based on the principles enumerated above is in operation in India for the last 14 years. The main objective of the programme of fertiliser trials in India has been to study the opitimum levels and combinations of the different fertilizers and the relative efficiency of diffeent forms of each fertilizer nutrients.

Out of a total of 318 districts in India, 184 districts which have either adequate rainfall or irrigation facilities, have been selected for the conduct of experiments. Each selected district has been divided into four agriculturally homogenous zones. In each zone, a centre consisting of about 100 sq. miles (about 100 villages) has been located at random each year for carrying out the experiments. There is a well laid out procedure for shifting these centres after two years of their operation. In each selected centre, three villages are selected at random for trials on each crop except legumes and in each selected village three fields are taken, one for each type of trial. Five categories of crops are covered, viz. summer and winter cereals, cash crops, oil seeds and legumes, the most important crops in each category being selected for the trials in a district.

In the experimental programme, there is provision for collecting soil samples from the experimental fields with a view to correlating the crop responses with the soil test data. Ancillary information, such as basal manuring given by the farmers, variety grown, soil type etc. are also being collected to find out the association, if any, between these factors and the responses to different treatments.

In addition to Simple Fertiliser Trials in cultivators' fields, Complex, Agronomic Experiments are underway at 73 Model Agronomic Experiment Centres in different soil regions of the centres. In the Simple Fertiliser Trials in Cultivators fields by irrigated in different soil class on paddy and wheat are given in tables I and II. An interesting fact revealed by these trials is the relatively good response to phosphate in many areas. Similarly, potash which was found not to show any response under the conditions at experimental stations has also shown fairly good response in many districts.

TABLE I

(Seth, 1967)

Response of rice to nitrogen, phosphate and potash according to soil classes
(Average for 1958-59 to 1961-62)
(Q/ha)

Soil Class	*No. of trials*	*Av. yield of untreated plots*	*Response to 22·4 Kg/ha of*		
			N	P_2O_5	K_2O
Alluvial	1601	11 1	2·4	1·4	0·9
Red	1595	15·8	2·6	2·0	1·2
Laterite	283	13·0	1·8	1·3	0·7
Red & Black	125	13·1	1·3	1·1	1·2
Coastal alluvium	74	16·0	3·2	1·9	1·6
Black	50	16·7	2·4	1·7	1·3
Deltaic alluvium	61	17·5	1·3	1.2	0·7
Submontane	40	11·6	2·2	2·0	1·6
Saline	28	11·3	2·2	1·0	0·3
Hilly	9	9·0	0·5	0·5	0·3

TABLE II

(Seth, 1967)

Response of Wheat (irrigated) to nitrogen, phosphate, and potash according to soil classes (Q/ha) (Average for 1958-59 to 1961-62)

Soil Class	*No. of trials*	*Av. yield of untreated plots*	*Response to 22·4 Kg/ha of*		
			N	P_2O_5	K_2O
Alluvial	2177	13·6	3·6	2·0	1·1
Saline	169	15·2	3·5	2·7	1·4
Red	164	8·7	2·6	1·8	1·3
Other Alluvial	95	16·5	3·5	1·7	0·8
Medium Black	47	9·1	1·4	1·3	0·8
Desert	44	12·4	3·3	1·6	0·6
Red & Black	42	9·9	2·6	1·5	1·1
Black	41	8·8	2·5	1·9	1·0
Grey brown	42	12·4	1·5	2·1	1·6
Laterite	36	11·3	3·3	1·7	0.9
Deltaic alluvium	19	14·2	2·0	3·3	3·1
Sub-montane	11	13·4	6·3	3·6	1·8

Table I shows that there is no appreciable variation in the response to nitrogen in the different soil classes. However, in red and black soils and in deltaic alluvium soils, the response to nitrogen was comparatively low. Response to phosphate was relatively higher in sub-montane soils of Punjab and Himachal Pradesh and red soils (about 2.0 Q/ha). Response to potash was higher in coastal alluvium and sub-montane soils.

Table II shows that with irrigated wheat, of the three soil classes, alluvial, saline and red, the response to N and P was less in red soils as compared to the other two types. The general conclusion has been made that all crops investigated responded well to the application of nitrogen.

India's agriculture has far long been a gamble for rains. It has, however, been established that there is considerable scope for improving Indian agriculture by proper management of soil and water and growing varieties suitable for varied weather conditions. While a good start has been made with the soil tests that would help formulation of appropriate fertiliser recommendations, follow up of the soil test recommendations, has been frequently inadequate. The bottlenecks in this respect need rectification. Further since 70 per cent of the country's cultivable land will continue to be rainfed for a long time to come, a full knowledge of the soil resources by soil survey and soil testing is urgently called for. This will lead to increased efficiency of the inputs like fertilisers and irrigation, etc. There is no doubt that scientific agricultural technology should make it possible to increase the production per unit area per unit time very considerable.

REFERENCES

1. Agarwal, R.R. and Mehrotra C.L. (1952, 1953, 1958) Soil Survey and Soil Work in Uttar Pradesh Department of Agriculture, U.P., 2 (1952), 3 (1953), 4 (1958).

2. Bains, S.S., Prasad, Rajendra and Bhatia, M.L. (1967). A summary of results obtained under the model agronomic experiment scheme (ICAR) during the period 1957-65. Paper presented at the workshop of Agronomists, Soil Scientists, Statisticians, Agricultural Engineers, 6-10 June, 1967.

3. Basu, J.K. and Sirur, S.S. (1938). Soils of Deccan Canals. I. Genetic soil survey and soil classification. Nira Right Bank and Pravara Canals. *Indian J. Agri. Sci.* **8**, 637-697.

4. Biswas, T.K. and Gowande, S.P. (1962). Studies in genesis of catenary soils on sedimentary formation in Chhatisgarh basin in Madhya Pradesh. *Jour. Indian Soc. Soil Sci.* **10**, 223-34.

5. Mukherjee, B.K. and Dass, N.K. (1940). Studies on Kumaon Hill Soils I. Soil survey at the Government Orchard, Chaubattia, Formation of genetic groups, *Indian J. Agri. Sci.* **10**, 990-1020.

6. Panse, V.G. and Abraham, T.P. (1960) Simple Scientific Experiments on Farmers land, 31 Session of the International Statistical Institute, Brussels.

7. National Institute of Sciences of India (1964). Symposium on Fertility of Indian Soils.

8. Raychaudhuri, S.P. (1941) Studies on Indian Red Soils III, General Morphological characteristics of some profiles, *Indian J. Agri. Sci.* **11**, 220-235.

9. Satyanarayana, K.V.S. and Thomas, P.K. (1961) Studies on latenite and associated soils, I. Field characteristics of Laterite of Malabar and South Kanara, *Indian Soc. Soil Sci.* **9**, 107-118.

10. Seth, G.R. (1967) Review of the progress of work under the coordinated scheme of simple fertiliser trials on cultivators fields during the scheme and third plan periods. Paper presented at the workshop of Agronomists, Soil Scientists, Statisticians, Agricultural Engineers, 6-10 June, 1967.

ASYMMETRIC ROTATION DESIGNS IN SAMPLING ON SUCCESSIVE OCCASIONS

By

V.B. SAVDASIA AND G.R. SETH

A population whose character changes over time may have to be repeatedly studied at intervals, say, every year or even at shorter periods. For example, milk production in a country not only changes from year to year but has seasonal fluctuations. It varies from summer to rainy and from rainy to winter and from winter to summer season. In such a case, estimates may be needed not only for annual milk production but also for each season to study the seasonal changes in a given year as well as the changes in milk production from the last year in a given season. These studies on several occasions have a special importance when planned efforts are being made to modify the character of the population, such as, the rate of economic growth in a country.

The designing of sampling investigations to be repeated on several occasions is similar to the one in which a number of characters of a given population are under study. It faces the same problems as encountered in obtaining optimum sampling designs for multiple characters. An optimum design for the study of one character may not be optimum for another character and some compromise design has to be developed which is optimum in a certain over-all sense. However, designing studies for a dynamic population introduce some new aspects, such as, whether the sampling units be identical at different points of time. If not, what proportion of the units should be retained and how one utilises the information from the past to improve the estimates for the current occasion.

This problem has been studied by a number of authors. Jessen[4] first investigated this problem for two occasions and his method has been extended to the study of a population for more than two occasions by Patterson[7]. Tikkiwal[11] obtained independently the results obtained by Patterson for a somewhat more general method of correlation between the character under study on different occasions. He also extended these results to the simultaneous study on several characters. A number of workers at the I.A.R S, New Delhi (see references) and elsewhere have studied this problem for multi-stage designs, as also using ratio estimates instead of regression estimates and varying probabilities instead of equal probabilities. Narain[6] and Tikkiwal have studied the efficiency of the regression estimate given by Patterson where the value of the regression coefficient is not known and is to be estimated from the sample. Eckler[1] has studied the estimates for a different type of design called by Wilks as 'rotation sampling' of different levels. *h-th* level of rotation means that n observations studied on the i-th occasion are also studied on all the previous occasions upto $i-h+1$th occasion. Hansen, Hurwitz and Madow[2] have used this technique to obtain estimates of total sale of retail store in which a particular pattern of a 12-month rotation was adopted.

Hansen *et-al*[3] in 'The Redesign of the census current population survey' have considered a design in which a rotation sampling design is imposed within each primary unit, mainly for the purpose of reducing response resistance and to reduce the within primary component of variance of the estimates under certain conditions.

Rao and Graham[8] have generalised the rotation pattern of Hansen *et.al*[3] for unistage sampling. Under this rotation pattern, a group of n_2 units stays in the sample for r occasions (n=sample size on any occasion$=n_2 r$), leaves the sample for m occasions, comes back into the sample for another r occasions, then leaves the sample for m occasions, and so on. The population size N is a multiple of n_2 and is finite. They have considered the case $m \geqslant r$. The composite linear estimate considered by them is of the same form as considered by Hansen *et al*[3]. The variance of the estimate has been obtained under the exponential correlation model and also under the arithmetic model. It is concluded that under the exponential model, (*i*) the optimum value of r is 2, (*ii*) the value of the variance of moderate m is virtually the same as that for $m=\infty$ and that (*iii*) as revealed by comparison of the variances under arithmetic model, moderate deviations from the exponential correlation model do not affect the efficiency of the estimates materially.

Studies of sampling designs for more than two occasions have been generally made under the assumption that the correlation between the i *th* and j *th* occasion is given by $\rho\ |i-j|$ where ρ is less than 1. In the present study, the correlation between any two occasions is assumed to be constant and the rotation pattern is an asymmetric one in contrast to the symmetric pattern study by Eckler, Hansen *et.al* and Rao and Graham.

General Rotation Pattern

In an h occasion asymmetric rotation pattern, the sample on the first occasion consists of h groups of units, the h-th of which is of size $n[1-(h-1)p](p<1)$ and is to be dropped from the sample after the first occasion, while the m th of the remaining $(h-1)$ groups consists of np units which are to be retained in the sample upto $(1+h-m)$th occasion. The sample on the i *th* occasion $(i \geqslant 2)$ consists of $(h+1)$ groups of the units, $(h+1)$th of which is of size $n(1-hp)$, while the remaining h groups are of size np each. The $(h+1)$*th* group enters the sample on the i *th* occasion and is to be observed on that occasion only. This group of units may be called the 'non-repeating' group. The m *th* of the remaining h groups which had entered the sample on $(i-m+1)$*th* occasion (first if $i<m$) and is to be retained in the sample upto $(i+h-m)$*th* occasion.

It can be seen that the maximum number of occasions for which any group of np units stays in the sample is h and that on any occasion i, $(i \geqslant h)$, the h-*th* group has stayed in the sample for the maximum number of occasions permissible *i.e.*, h.

The schematic pattern upto 5 occasions for $h=3$ is presented below :

$$\begin{array}{llllll}
\left[X^{1}_{1,1}\right] & \left[X^{1}_{1,2}\right] & \left[X^{1}_{1,3}\right] & & & \\
\left[X^{2}_{2,2}\right] & \left[X^{2}_{1,2}\right] & \left[X^{2}_{1,3}\right] & \left[X^{2}_{2,4}\right] & & \\
\left[X^{3}_{3,3}\right] & - & \left[X^{3}_{1,3}\right] & \left[X^{3}_{2,4}\right] & \left[X^{3}_{3,5}\right] & \\
\left[X^{4}_{4,4}\right] & - & - & \left[X^{4}_{2,4}\right] & \left[X^{4}_{3,5}\right] & \left[X^{4}_{4,6}\right] \\
\left[X^{5}_{5,5}\right] & - & - & - & \left[X^{5}_{3,5}\right] & \left[X^{5}_{5,7}\right]
\end{array}$$

Here $\left[X^{k}_{j,i}\right]$ denotes the group of units observed on the k-th occasion, entering the sample on the jth occasion and retained in the sample upto i-th occasion. For $j \neq i$, the number of observations in $X^{k}_{j,i}$ is np and for $j=i$, the size will be $n(1-hp)$ except when $j=i=1$ in which case the size will be $n[1-(h-1)p]$.

It can be observed that the number of units common between the i-th and the j-th occasions $(i>j)$ is $n(h-i+j\ p$ if $(i-j) \leqslant h-1$ and is zero if $(i-j) \geqslant h$. Thus the number of units common to any two consecutive occasions is $n(h-1)p$.

In practice, the selection of h in a large scale survey is determined by the seasonality of the character or the periodicity with which the estimates of population mean or the changes in the population mean are studied. In case of assessment of achievements of five year plan schemes h may be equal to 5, whereas studying the seasonal changes in milk yields h can be 3 or 4 depending upon whether a year is divided into three or four seasons.

Unlike the symmetric rotation pattern of Rao and Graham (8) where all the groups into which a sample is divided are of equal size and each of the groups of units is observed uniformly for the rotation period of h occasions, it may be seen that the rotation period studied here has a non-repeatable group of units of a size different from that of a repeatable group of units and certain groups of units observed on any occasion upto $h-1$ are to be studied for a total number of occasions less than h.

Let us assume that the population under study is infinite and that corresponding to each unit on the population, there is a finite value of the character X under study on the i-th occasion denoted by x^{i}. We further assume that the population variance (σ^{2}) is finite and remains invariant over occasions. Further the correlation coefficient between the values of the character X on any two occasions does not depend upon the occasions and has the same value for all pairs of occasions. Such a pattern of correlation has been observed on the area under arecanut in annual surveys conducted by the Indian Council of Agricultural Research in Assam. The sample size is assumed to be constant for the various occasions. Pattern of rotation sampling assumed is the h-occasion asymmetric rotation pattern described in an earlier paragraph in which a unit once observed is repeated on successive occasions upto a maximum of h occasions. Further let $\bar{X}^{i}_{j,k}$ denote the mean of the n' observed on the ith occasion which were first observed on the j-th occasion and are retained in the samples upto k-th occasion. Correlation between $\bar{X}^{i}_{j,k}$ and $\bar{X}^{i'}_{j,k}$ is given by $\rho\sigma^2/n'$ for i, i' lying between j and k where ρ is the correlation coefficient between any two occasions.

We consider estimates of the population mean on any occasion i based on all the sample observations made (*i*) upto and including the ith occasion (*ii*) upto and including i'th occasion ($i' > i$). Also estimates of the change in the value of the population mean between ith and i'th occasions are studied. We first consider best linear unbiased estimates.

The following two results obtained by Patterson and Eckler will be extensively used.

1. θ^* is the best linear unbiased estimate of $\alpha = \sum_{i=1}^{m} c_i \alpha_i$ based on observations upto and including mth occasion where c_i are known constants and α_i is the unknown population mean on the i-th occasion, if and only if Cov $(\theta^*, x_{ij}) = Z_{i\alpha}$ for all combinations of i and j where x_{ij} is the value of the jth unit observed on the ith occasion. This leads us to the conclusion that

$$\text{Cov}(\bar{X}_i, \theta^*) = \lambda_{i\alpha}$$

and $$\text{Var}(\theta_i^*) = \text{Cov}(X_{ij}, \theta_i^*) \text{ where}$$

θ_i^* is the Blue for α_i and $\bar{X}_i$ is the mean of the values of the units observed on the ith occasion.

2. If P is the finite incomplete matrix $[X_{ij}]$ formed by the values of the distinct units observed upto and including mth occasion, occasion i forming the rows and the units forming the columns ; $[w_{ij}]$ is the corresponding matrix of weights for minimum variance unbiased linear estimate $\sum w_{ij} x_{ij}$; P_u $(r \times c)$ is any one of the complete rectangular submatrices into which P can be decomposed and $W_u(r \times c)$ the corresponding complete rectangular matrix for similar decomposition of $W = [w_{ij}]$, then w_{ij}'s in any row of $W_u(r \times c)$ are constant from one column to another *i.e.*

$$\begin{aligned} w_{11} &= w_{12} = \ldots\ldots = w_{1c} \\ w_{21} &= w_{22} = \ldots\ldots = w_{2c} \\ \cdots \quad &\cdots \quad \cdots \quad \cdots \quad \cdots \\ w_{r1} &= w_{r2} = \ldots\ldots = w_{rc}. \end{aligned}$$

Estimates of population mean.

It is assumed that ρ and σ^2 are known. In case they are not known, these may be calculated from the sample observations and to that extent the efficiency of the estimates is effected.

1. Best linear unbiased estimate (BLUE).

$h=2$

The BLUE for $h=2$ will be of the following form :

$$\hat{\bar{X}}_i=(-a_i-b_i-c_i)\,\hat{\bar{X}}_{i-1}+(1-d_i)\left[\frac{(1-2p)\bar{X}^{i}_{i,\,i}+p\,\bar{X}^{i}_{i,\,i+1}}{1-p}\right]$$

$$+d_i\bar{X}^{i}_{i=1,\,i}+a_i\bar{X}^{i-1}_{i-1,\,i}+b_i\,\bar{X}^{i-1}_{i-2,\,i-1}+c_i\bar{X}^{i-1}_{i-1,\,i-1}$$

where $\bar{X}_i$ represents the BLUE based on observation upto and including the ith occasion ($i\geqslant 2$).

There are four unknown coefficients and there will be only four independent covariance conditions, of which one is supplied by the ($i-2$)th occasion, two by ($i-1$)th occasion and one by ith occasion.

The solution of these four equations from covariance conditions give

$$b_i=c_i=0\ ;\ a_i=-d_i\rho$$

and

$$d_i=\frac{p}{1-(1-p)\rho^2+p(1-p)V_{i-1}\rho^2}\text{ with }d_1=p$$

where

$$V_{i-1}=V(\hat{\bar{X}}_{i-1})n/\sigma^2.$$

Therefore, $\bar{X}_i$ can be put as

$$\hat{\bar{X}}_i=d_i\left[\bar{X}^{i}_{i-1,i}+\rho\left(\hat{\bar{X}}_{i-1}-\bar{X}^{i-1}_{i-1,\,i}\right)\right]$$

$$+(1-d_i)\left[\frac{(1-2p)\ \bar{X}^{i}_{i,\,i}+p\,\bar{X}^{i}_{i,\,i+1}}{1-p}\right]$$

This is the same, as it should be, as the estimate obtained by Patterson[3], under the appropriate exponential correlation model, which makes all partial regression coefficients, two occasions more than one apart, zero,

The variance $V(\hat{\bar{X}}_i)=V_i\sigma^2/n$, where V_i is given by

$$V_i=\frac{1-\rho^2+p\rho^2V_{i-1}}{1-\rho^2+p\rho^2(1+V_{i-1})-p^2\rho^2V_{i-1}}$$

The limiting value of V_i is the same as that obtained by Patterson.

Case when $h=3$

The problem of finding Blue in case of three occasions rotation design is somewhat difficult. The considerations to be borne in mind while building up an iterative estimate are that the number of unknown coefficients is equal to the number of independent covariance conditions and that the set of simultaneous equations obtained from covariance conditions should have a non-trivial solution (the unknown coefficients should be all zero). The rotation pattern, in the present case, suggests an estimate of following form :

$$\hat{\bar{X}}_i=(1-E_i-F_i)\left[\frac{(1-3p)\,\bar{X}^{i}_{i,\,i}+p\bar{X}^{i}_{i,\,i+2}}{1-2p}\right]+E_i\bar{X}^{i}_{i-2,\,i}$$

$$+F_i\bar{X}^{i}_{i-1,\,i+1}-(H_i+I_i+J_i)\hat{\bar{X}}_{i-1}+H_i\bar{X}^{i-1}_{i-2,\,i}+I_i\bar{X}^{i-1}_{i-1,\,i+1}$$

$$+J_i\bar{X}^{i-1}_{i-1,\,i-1}-M_i\hat{\bar{X}}_{i-2}+M_i\bar{X}^{i-2}_{i-2,\,i}$$

The unknown coefficients are to be determined by use of the covariance conditions. The covariance conditions are obtained by considering the group of observations on the *i*th, $(i-1)$th and $(i-2)$th occasions, remembering that under the present correlation model,

$$\text{Cov}\left(\bar{X}^{i}_{i-2,\,i},\ \hat{\bar{X}}_{i-2}\right)=\text{Cov}\left(\bar{X}^{i-1}_{i-2,\,i},\ \hat{\bar{X}}_{i-2}\right)=\rho V(\hat{\bar{X}}_{i-2}).$$

The covariance conditions thus obtained from various occasions are as follows :

(1) From the *i*th occasion :

$$(1-E_i-F_i)\sigma^2/n(1-2p)=-(H_i+I_i+J_i)\text{Cov}\left(\bar{X}^{i}_{i-2,\,i},\ \hat{\bar{X}}_{i-1}\right)-M_i\rho V(\hat{\bar{X}}_{i-2})$$

$$+(E_i+H_i\rho+\text{M}_i\rho)\sigma^3pn$$

$$=-(H_i+I_i+J_i)\rho V(\bar{X}_{i-1})+(F_i+I_i\rho)\sigma^2/np$$

$$=V(\hat{\bar{X}}_i)$$

(2) From the $(i-1)$th occasion :

$$J_i\sigma^2/n(1-3p)=(I_i+F_i\rho)\sigma^2/np$$

$$=(E_i\rho+H_i+M_i\rho)\sigma^2/np-M_i\rho V(\hat{\bar{X}}_{i-2})$$

$$=-M_i\,\text{Cov}\left(\hat{\bar{X}}_{i-2},\ \bar{X}^{i-1}_{i-3,\,i-1}\right)$$

(3) From the $(i-2)$th occasion :

$$(E_i\rho+H_i\rho+M_i)\sigma^2/np=0$$

Covariance conditions from the earlier occasions do not provide any further information about the coefficients. It can be seen that there are six unknown coefficients to be determined and we have got six simultaneous linear equations. The solution of the above six simultaneous equations give

$$E_i=\frac{p[1-\rho^2+p\rho(\alpha-\beta)][1-\rho^2-p\rho(\gamma-\rho V_{i-2})}{[p[1-\rho^2+p\rho(\alpha-\beta)][1-\rho^2-p\rho(\gamma-\rho V_{i-1})]}$$

$$+(1-2p{,}p\rho V_{i-1}[1-\rho^2-p\rho(\gamma-\rho V_{i-1})][\rho-\rho^2$$

$$+p\rho(\alpha-\beta)-(1-3p)\rho(1-\rho)\beta]$$

$$+(1-2p)\rho(\rho-p\rho V_{i-1})[(1-\rho)p\beta-\gamma]$$

$$+(1-p)[\gamma-\rho^2(1-\rho)p\beta-p^2\rho(1-\rho)\beta(\gamma-\rho V_{i-1}]$$

$$M_i=\frac{-E_i\rho(1-\rho)}{[1-\rho^2+p\rho(\alpha-\beta)]}$$

$$H_i=\frac{-E_i\rho[1-\rho+p(\alpha-\beta)]}{[1-\rho^2+p\rho(\alpha-\beta)]}$$

$$J_i=\frac{E_i\rho(1-3p)(1-\rho)\beta}{[1-\rho^3+p\rho(\alpha-\beta)]}$$

$$I_i=\frac{E_i\rho[(1-\rho)p\beta-\gamma)]}{[1-\rho^2+p\rho(\alpha-\beta)][1-\rho^2-p\rho(\gamma-\rho V_{i-1})]}$$

$$F_i=\frac{E_i[\gamma-\rho(1-\rho)p\beta\{(\rho+p(\gamma-\rho V_{i-4})\}]}{[1-\rho^2+p\rho(\alpha-\beta)][1-\rho^2-p\rho(\gamma-\rho V_{i-1})]}$$

where
$$\alpha=\text{Cov}\left(\bar{X}^{\,i-1}_{i-2,\,i},\ \hat{\bar{X}}_{i-2}\right)=\rho V_{i-2}\sigma^2/n$$

$$\beta=\text{Cov}\left(\bar{X}^{\,i-1}_{i-3,\,i-1},\ \hat{\bar{X}}_{i-2}\right)=(F_{i-2}+I_{i-2})\rho\sigma^2/np-(H_{i-2}$$

$$+I_{i-2}+J_{i-3})\rho V(\hat{\bar{X}}_{i-3})$$

$$\gamma=\text{Cov}\ (\bar{X}^i{}_{i-2,\,i},\ \hat{\bar{X}}_{i-1})=(F_{i-1}+I_{i-1})\rho\sigma^2/np-(H_{i-1}+I_{i-1}+J_{i-1})\rho V(\hat{\bar{X}}_{i-2})$$

$$V_i=V(\hat{\bar{X}}_i)n/\sigma^2$$

$$v=\rho p(\gamma-\rho V_{i-1})[1-\rho+p(\alpha-\beta)-(1-3p)(1-\rho)\beta]$$
$$+(1-3\rho^2)+2\rho^3+(1-\rho)p\rho+(2\alpha-\beta)$$

As nontrivial solution is obtained, we conclude that $\hat{\bar{X}}_i$ is the minimum variance estimate of population mean on the i-th occasion. The variance $V(\hat{\bar{X}}_i)$ is given by

$$V(\hat{\bar{X}}_i)=\frac{(1-E_i-F_i)\sigma^2}{n(1-2p)}$$

For the values of $h>3$, both the form of the estimate and the calculation of w_{ij}'s becomes complicated.

The form of $\hat{\bar{X}}_i$ for $h=4$ and $h=5$ are as follows :
Case when $h=4$

To obtain a BLUE, we start with a general form given below :

$$\hat{\bar{X}}_i=(1-a'_i-b'_i-c'_i)\,[(1-4p)\bar{X}^i_{i,i}+p\bar{X}^i_{i,\,i+3}]/(1-3p)$$
$$+a'_i\bar{X}^i_{i-3,\,i}+b'_i\bar{X}^i_{i-2,\,i+1}+c'_i\bar{X}^i_{i-1,\,i+2}-(d'_i+e'_i+f'_i+g'_i+h'_i)\hat{\bar{X}}_{i-1}$$
$$+d'_i\bar{X}^{i-1}_{i-4,\,i-1}+e'_i\bar{X}^{i-1}_{i-3,\,i}+f'_i\bar{X}^{i-1}_{i-2,\,i+1}+g'_i\bar{X}^{i-1}_{i-1,\,i+2}$$
$$+h'_i\bar{X}^{i-1}_{i-1,\,i-1}-\left(i'_i+j'_i+k'_i+l'_i+m'_i\right)\hat{\bar{X}}_{i-2}+I'_i\bar{X}^{i-2}_{i-5,\,i-2}+j'_i\bar{X}^{i-2}_{i-4,\,i-1}$$
$$+k'_i\bar{X}^{i-2}_{i-3,\,i}+l'_i\bar{X}^{i-2}_{1-2,\,i+1}+m'_i\bar{X}^{i-2}_{i-2,\,i-2}+s'_i\bar{X}^{i-3}_{i-3,\,i}-s\hat{\bar{X}}_{i-3}$$

Case when $h=5$

The form of the BLUE obtained in case of $h=3$ and 4 suggest the following form of the BLUE when $h=5$:

$$\hat{\bar{X}}_i=\left(1-a''_i b''_i-c''_i-d''_i\right)\left[(1-5p)\ \bar{X}^i_{i,\,i}+p\bar{X}^i_{i,\,i+4}\right]\Big/(1-4p)$$
$$+a''_i\bar{X}^i_{i-4,\,i}+b''_i\bar{X}^i_{i-3,\,i+1}+c''_i\bar{X}^i_{i-2,\,i+2}+d''_i\bar{X}^i_{i-1,\,i+3}$$
$$-\left(e''_i+f''_i+g''_i+h''_i+i''_i+j''_i\right)\hat{\bar{X}}_{i-1}+e''_i\bar{X}^{i-1}_{i-5,\,i-1}$$
$$+f''_i\bar{X}^{i-1}_{i-4,\,i}+g''_i\bar{X}^{i-1}_{i-3,\,i+1}+h''_i\bar{X}^{i-1}_{i-2,\,i+2}+i''_i\bar{X}^{i-1}_{i-1,\,i+3}$$
$$+j''_i\bar{X}^{i-1}_{i-1,\,i-1}-\left(k''_i+l''_i+m''_i+n''_i+p''_i+q''_i\right)\hat{\bar{X}}_{i-2}$$
$$+k''_i\bar{X}^{i-2}_{i-6,\,i-2}+l''_i\bar{X}^{i-2}_{i-5,\,i-1}+m''_i\bar{X}^{i-2}_{i-4,\,i}+n''_i\bar{X}^{i-2}_{i-3,\,i+1}$$
$$+p''_i\bar{X}^{i-2}_{i-2,\,i+2}+q''_i\bar{X}^{i-2}_{i-2,\,i-2}-\left(s''_i+t''_i+u''_i+v''_i+w''_i\right)\hat{\bar{X}}_{i-3}$$
$$+s''_i\bar{X}^{i-3}_{i-6,\,i-2}+t''_i\bar{X}^{i-3}_{i-5,\,i-1}+u''_i\bar{X}^{i-3}_{i-4,\,i}+v''_i\bar{X}^{i-3}_{i-3,\,i+1}$$
$$+w''_i\bar{X}^{i-3}_{i-3,\,i-3}-r''_i\hat{\bar{X}}_{i-4}+r''_i\bar{X}^{i-4}_{i-4,\,i}.$$

The above results hold good for $i \geqslant h$. For $i < h$ say j, the form of the Blue will be the same as that of an estimate corresponding to j-asymmetric rotation pattern.

As the Blue estimate becomes more and more complicated for higher values of h alternate estimates are therefore, suggested which are comparatively easy to calculate though somewhat less efficient than Blue.

Combined regression estimate (CRE):

As some of the units observed on i-th occasion have been observed on earlier occasions also, we can use the regression technique to develop an estimate of the population mean on the i-th occasion, which takes advantage of the information on earlier occasions. As the matching scheme is such that a fixed number of observations made on i th occasion are also measured on earlier occasions upto $(i-t)(t=1, 2,\ldots h-1)$, we can build up a multiple regression estimate for each of the matching groups of observations and in addition, we can build up an independent estimate based on $n\{1-(h-1)p\}$ units entering the sample on the i th occasion. By combining these h estimates linearly, we get a combined regression estimate $\hat{X}_i'''$ (CRE) for the i-th occasion ($i \geqslant h$) given by

$$\sum_{t=1}^{h-1} A_t\Big[\overline{X}^{\,i}_{i-t,\, i-t+h-1} + \sum_{j=1}^{t} \beta_{i,\, j-1(t-1)} \Big(\hat{\overline{X}}'''_{i-1} - \overline{X}^{\,i-1}_{i-t,\, i-t+h-1}\Big)$$

$$+ A_h\Big[\Big(1-hp\Big)\overline{X}^{\,i}_{i,\, i} + p\,\overline{X}^{\,i}_{i,\, i+h-1} \;/1-(h-1)p \qquad \ldots(2)$$

where $\hat{\overline{X}}'''_{i-1}$ is CRE for the $(i-1)$th occasion, and where $\beta_{i,\, i-j.(t-1)}$ is the partial regression coefficient between i-th and $(i-j)$th occasion fixing the values on other occasions between i and $i-t$ is given by $\rho/(1+(t-1)\rho)$.

The constants A's have to be determined, subject to their sum being unity. such that the resulting estimate has a minimum variance among its class of estimates.

Simplified combined regression estimate (SCRE) :

The expression for variance of CRE as given in (2) is difficult to compute. A simpler estimate obtained by replacing $\hat{\overline{X}}'''_{i-j}$ by $\overline{X}_{i-j}$, the mean of all the observations on the $(i-j)$th occasion, in the expression (2) and the estimate may be denoted by $\hat{\overline{X}}'_i$ and called the simplified combined regression estimate (SCRE), as its variance is easier to compute as compared to that of CRE. Expression for the variance of $\hat{\overline{X}}'_i$ for $h=2$ and 3 are as below :

When $h=2$:

$$V\Big(\hat{\overline{X}}_i\Big) = \frac{(1-\rho^2+p\rho^2)}{(1-\rho^2+2p\rho^2-p^2\rho^2)} \;\frac{\sigma^2}{n}$$

When $h=3$:

$$V\left(\hat{\bar{X}}'_i\right)=\frac{[(1+2\rho)(1-\rho^2)^2+p(3\rho^2+2\rho^3)(1-\rho^2)+p^2(4\rho^3+\rho^4-4\rho^5)-4p^4\rho^6]\sigma^2/n}{[(1+2\rho,(1-p^2)^2+6p\rho^2(1-\rho^2)(1+\rho)-p^2(5p^3-8p^4)-p^3(8\rho^3+6\rho^4-8\rho^5)-4p^4\rho^6+8p^5\rho^6]}$$

Composite estimate (CE) :

This estimate is the one considerrd by Hansen et al[3] and is given by

$$\hat{\bar{X}}''_i=a_i\left[\hat{\bar{X}}''_{i-1}+\bar{X}^{i}_{(i-1,\,i)}-\bar{X}^{i-1}_{(i-1,\,i)}\right]+(1-a_i)\bar{X}_i$$

where $\bar{X}^{i}_{(i-1,\,i)}$ and $\bar{X}^{i-1}_{(i-1,\,i)}$ are the means of the common observations between $(i-1)$th and i-th occasion.

Variances of CE and SCRE are tabulated below for $\rho=\cdot 9$ and $h=3$.

Variance of CE, SCRE and BLUE when $\rho=0{\cdot}9$, $h=3$

$\left(\text{in multiples of } \frac{\sigma^2}{n}\right)$

	Occasion	*CE*	*SCRE*	*BLUE*
$p=0{\cdot}1$	1	1·0000	1·0000	1·0000
	2	0·7858	0·7309	0·7309
	3	0·7579	0·6595	0·6533
	4	0·7506	0·6595	0·6380
	5	0·7492	0·6595	0·6311
$p=0{\cdot}2$	1	1·0000	1·0000	1·0000
	2	0·8106	0·7256	0·7256
	3	0·7622	0·6422	0·6178
	4	0·7396	0·6422	0·5905
	5	0·7313	0·6422	0·5655
$p=0{\cdot}3$	1	1·0000	1·0000	1·0000
	2	0·8695	0 7767	0·7767
	3	0·8387	0·6997	0 6023
	4	0·8171	0·6997	0·6085
	5	0·8113	0·6997	0·5803

It may be seen that variance of SCRE lies between those of CE and BLUE, and SCRE becomes gradually inefficient compared to BLUE with increasing values of *p*, the replacement fraction.

Estimate of change in the population mean between i th and j-th occasion $(j>i)$.

An estimate of change in the population mean (α) between *j-th* and *i th* occasions is given by $\hat{\bar{X}}_j - \hat{\bar{X}}_i$ $(j>i)$. But this is not the best estimate of change in the population mean. BLUE for estimating $\alpha_j - \alpha_i$ has to be based on all the observations included upto the *j*-th occasion and will be given by $\hat{\bar{X}}_j - {}_j\hat{\bar{X}}_i$, where e ${}_j\hat{\bar{X}}_i$ is the Blue for the occasion $i (i<j)$ when the information upto j th occasion is utilised. For this purpose we shall have to develop from for ${}_j\hat{\bar{X}}_j$.

Improved Blue for *i*-th occasion when information on subsequent occasion is available, we consider for simplicity the case for $h=2$ and $h=3$.

For $h=2$ we have

$$_{(i+1)}\hat{\bar{X}}_i = \hat{\bar{X}}_i - {}_{wi}\hat{\bar{X}}_{i+1} + w_i \left[\frac{(1-2p)\,\bar{X}^{i+1}_{i+1,\,i+1} + p\,\bar{X}^{i+1}_{i+1.\,i+2}}{1-p} \right]$$

where $\hat{\bar{X}}_i$ is the Blue based on observations upto the *i*-th occasion.

Applying the necessary covariance conditions to be satisfied by a BLUE, for $(i+1)$th occasion, we get

$$w_i = V_i \rho (1-p)$$

where $V_i = V(\hat{\bar{X}}_i)\, n/\sigma^2$.

Therefore,

$$_{i+1}\hat{\bar{X}}_i = \hat{\bar{X}}_i - V_i(1-p)p \left[\hat{\bar{X}}_{i+1} - 1/(1-p) \left((1-2p)\bar{X}^{i+1}_{i+1,\,i+1} + p\bar{X}^{i+1}_{i+1,\,i+2} \right) \right]$$

Observing that cov $(\hat{\bar{X}}_i, \hat{\bar{X}}_{i+1}) = d_{i+1}(1-d_i)\sigma^2\rho/n(1-p)$ the variance of the estimate is given by

$$V({}_{i+1}\hat{\bar{X}}_i) = \frac{(1-d_i)\sigma^2}{n(1-p)} \left[1-\rho^2(i-d_i)d_{i+1}\right]$$

where d_i is the coefficient of $\rho\,\hat{\bar{X}}_{i-1}$ in the expression of $\hat{\bar{X}}_i$.

The covariance between $\hat{\bar{X}}_{i+1}$ and ${}_{i+1}\hat{\bar{X}}_i$ is given by

$$\text{Cov}(\hat{\bar{X}}_{i+1}, {}_{i+1}\hat{\bar{X}}_i) = \text{Cov}(\hat{\bar{X}}_i, \hat{\bar{X}}_{i+1}) = d_{i+1}(1-d_i)\sigma^2\rho/n(1-p)$$

when information upto $(i+k)$th occasion is available, the improved Blue for *i*-th occasion is give by

$$_{i+k}\hat{\bar{X}}_i = {}_{i+k-1}\hat{\bar{X}}_i - q_{i+k}\hat{\bar{X}}_{i+k} + q_{i+k}\left[(1-2p)\bar{X}^{i+k}_{i+k,\,i+k} + p\bar{X}^{i+k}_{i+k,\,i+k+1} \right] / (1-p)$$

where the weights q_{i+k} obtained by applying the covariance conditions for $(i+k)$th occasion are given by

$$q_{i+k}=p^k(1-d_i)d_{i+1}\,d_{i+2}\,\cdots\,d_{i+k-1}$$

The results obtained above are the same as obtained by Patterson[3] under exponential correlation model.

h=3

The improved Blue for the ith occasion when the information up to j-th occasion $(j>i)$ is available can be found by the use of recurrence formula given by

$$\hat{{}_jX_i}={}_{j-1}\hat{\bar{X}}_i-B_i\rho\left(\hat{\bar{X}}_{j-1}-\hat{\bar{X}}^{j-1}_{j-1,\,j+1}\right)+(\alpha_i+\gamma_i)\ \hat{\bar{X}}_j-\alpha_1\bar{X}^{j}_{j-1,\,j+1}$$

$$-\gamma_i\left[(1-3p)\bar{X}^{j}_{j,j}+p\,\bar{X}^{j}_{j,\,j+2}\right]/(1-2p)$$

Necessary covariance conditions to be satisfied by ${}_j\hat{\bar{X}}_i$ give

$$\xi-\alpha_i\rho^2V(\hat{\bar{X}}_{j-1})-\alpha_i\sigma^2(1-\rho^2)/np$$

$$=\theta-\alpha_i\rho\eta$$

$$=\gamma_i\sigma^2/n(1-2p)$$

where $\xi=\operatorname{cov}\left(\bar{X}^{j}_{j-1,\,j+1},\ {}_{j-1}\hat{\bar{X}}_i\right)=(\alpha_{i-1}+\gamma_{i-1})\rho V(\hat{\bar{X}}_{j-1})-\gamma_{i-1}\sigma^2/n(1-2p)$

$$\theta=\operatorname{cov}\left(\bar{X}^{j}_{j-2,j},\ {}_{j-1}\hat{\bar{X}}_i\right)$$

$$=(\alpha_{i-2}+\gamma_{i-2})\rho V(\hat{\bar{X}}_{j-2})m-\gamma_{i-2}\sigma^2/n(1-2p)$$

$$-\alpha_{i-1}\rho[V(\hat{\bar{X}}_{j-2})+\sigma^2(1-\rho)/np]+(\alpha_{i-1}+\gamma_{i-1})\eta$$

$$\eta=\operatorname{cov}\left(\bar{X}^{j}_{j-2,\,j},\ \hat{\bar{X}}_{j-1}\right)=(F_{j-1}+I_{j-1})\rho\sigma^2/np-\psi_{j-1}\rho V(\hat{X}_{,-2})$$

$$\alpha_i=(\xi-\theta)/[V(\hat{\bar{X}}_{j-1})\rho^2+\sigma^2(1-\rho^2)/np-\rho\eta]$$

and $\gamma_i\sigma^2/n(1-2p)=(\alpha_i\rho\eta-\theta)$.

Estimate of the change of the population is then given by $\hat{\bar{X}}_j-{}_j\bar{X}_j\cdot\hat{\bar{X}}_i$.

Similarly estimate of change can be obtained, when CE and SCRE estimates are used. We may call the latter ISCRE. Reduction in variance of SCRE by using information on subsequent occasions is illustrated by the table :

Varianc of SCRE and ISCRE for h=5$\left(\text{in multiples of } \frac{\sigma^2}{n}\right)$

p \ ρ	0·9		0·8		0·7		0·6		0·5	
	SCRE	ISCRE	SCRE	ISCRE	SCRE	ISCRE	SCRE	ISCRE	SCRE	ISCRE
0·10	0·5630	0·4242	0·6718	0·5499	0·7563	0·6519	0·8172	0·7353	0·8768	0·8076
0·125	0·5841	0·4471	0·6759	0·5496	0·7519	0·6403	0·8125	0·7213	0·8688	0·7934
0·1667	0·6509	0·5015	0·7164	0·5826	0·7745	0·6572	0·8220	0·7272	0·8728	0·7927
0·20	0·7322	0·5653	0·7767	0·5977	0·8176	0·6952	0·8502	0·7713	0·8917	0·8115

It can be seen from the Table that the variance of SCRE of the population mean on the *i*-th occasion ($i \geqslant h$) for $h=3$, $\rho=\cdot 9$ and $p=\cdot 10$, the variance of the estimateis reduced from 0·5630 σ^2/n to 0·4242 σ^2/n.

Optimum p for a given h :

We discuss that optimum choice of replacement fraction for the estimate SCRE. For $h=2$, the optimum p is the same as obtained by Patterson. For $h=3$ and 4 optimum p's were obtained by solving 8th degree and 12th degree equation in p for specified values of ρ. For $h=5$, general expression for V_5 becomes complicated and variances were calculated for various values of p and ρ and a rough idea of the optimum p obtained from that Table. Optimum values of p for $h=3$, 4 and 5 are given as below :

Optimum p

ρ	$h=3$	$h=4$	$h=5$
0·9	0·1629	0·1118	0·08
0·8	0·2020	0·1394	0·10
0·7	0·2264	0·1568	0·125
0·6	0·2440	0·1696	0·125
0·5	0·2578	0·1797	0·125

It may be seen that optimum value p decreases with increasing h as well as ρ. Number of common observations given by $(h-1)pn$ for optimum p, however, remains fairly constant for various values of h. For $\rho=0{\cdot}5$, this value is approximately 1/2 and decreases to 1/3 for $\rho=0{\cdot}9$.

Optimum values of h for a given p have been studied though not reported here for lack of space. A pattern with higher values of h is optimum with low values of p. Values of h may be decreased as p is increased.

The role of non-repeating group of units has also been studied on an empirical basis for $h=3$, 4 and 5, for the case when SCRE is used. Broad conclusions are that the non-repeating group with p near the optimum value brings about an appreciable reduction in the variance under the present correlation model. Secondly, the variance of SCRE for $h=3$, 4, 5 differ very little from each other when the non-repeating group is absent, and thirdly that for the same sample size and for the same sample size of non-reapeating group, pattern with higher h give, smaller variance of SCRE than patterns with lower h. It means that it is advantageous to keep more number of smaller subgroups for a comparatively longer period than keeping a smaller number of larger subgroups for a shorter period.

Summary

The paper discusses the estimates on successive occasions when units are partially retained over various occasions. An asymmetric pattern of replacement is studied here for the correlation model where the correlation between units on any two occasions does not depend upon the occasions.

REFERENCES

1. Eckler, R.A. (1955) "Rotation Sampling", Ann. Math. Stat., 26.
2. Hansen, M.H. Hurwitz, W.N. and Madow W.G. (1953), Sample Survey Methods and Theory, Vol. 2. New York ; John Willey and Sons ; Inc.
3. Hansen, M.H. Hurwitz, W.N., Nisselson, H., and Steinberg, J. 1955). "The Redesign of the Census Current Population Survey", J. Am. Stat. Ass., 50.
4. Jessen, R.J. (1942) Statistical investigation of a sample survey for obtaining farm facts, Iowa Agri. Exp. Stat. Res. Bull., 304.
5. Kathuria, O.P. (1959) "Some Aspects of Successive Sampling in multistage designs" (unpublished thesis submitted towards fulfilment of requirements of Diploma I.C.A.R. New Delhi.)
6. Narain, R.D. (1953) "On the Recurrence Formula in Sampling on Successive occasions" J. Ind. Soc. Agri. Stat. 5.
7. Patterson, H.D. (1950) "Sampling on Successive Occasions with partial replacement of units", J. Roy. Stat. Soc., Series B, 12.
8. Rao, J.N.K. and Graham, J.E. "Rotation Design for Sampling on Repeated occasions" Technical Report 2.1, Iowa State University, Ames Iowa.
9. Seshavadhanulu, M. (1963) "Some contributions to the Theory of Sampling" (unpublished thesis submitted towards fulfilment of the requirements for Diploma, I.C.A.R., New Delhi.
10. Singh, B.D. (1962) "Use of Double Sampling in Repeated Surveys" (unpublished Thesis submitted towards fulfilment of requirements for Diploma I.C.A.R., New Delhi.
11. Tikkiwal, B.D. (1951) "Theory of Successive Sampling", (Unpublished thesis submitted towards fulfilments of requirements for Diploma I.C.A.R., New Delhi.)
12. Tikkiwal, B.D. (1964) "Theory of Successive Multiphase Sampling" Abstract, Ann. Math., Stat. 25.
13. Tikkiwal, B.D. (1955) "On the Efficiency of Estimates in Successive Multiphase Sampling" Abstract, Ann. Math. Stat., 26.
14. Tikkiwal, B.D. (1955) "Multiphase Sampling on Successive Occasions" Ph.D. Thesis, Faculty of North Carolina State College, Releigh, N.C., U.S.A.
15. Tikkiwal, B.D. (1966) 'A Further Contribution to the Theory of Unistage Sampling on Successive Occasions" H.J. Ind. Soc. Agri. Stat. 8.
16. Tikkiwal, B.D. (1956) "An Application of the Theory of Multiphase Sampling on Successive Occasions to Survey of Livestock Marketing" J. Karnatak Uni. i.
17. Tikkiwal, B.D. (1958) "Theory of Successive Two State Sampling", Abstract, Ann. Math. Stat., 29.
18. Tikkiwal, B D. (1964) "A note on Two State Sampling on Successive Occasions" Sankhya, Series A, 26.

THE GENETIC CONSTRUCTION OF HIGH YIELDING-CUM-HIGH QUALITY VARIETIES IN CEREALS

By

M.S. SWAMINATHAN*

The pattern of improvement in crop production in most countries reveals a discontinuous rise, associated with some major advance in crop husbandry. For example, the yield of wheat in the United Kingdom during the period 1400 to 1750, remained at the level of about 7·5 bushels per acre. In about 1750, the principle of crop rotation was introduced into agriculture and this resulted in the per-acre yield increasing from about 7·5 to 20 bushels, The beneficial effect of crop rotation on yield is seen in other countries also. The highest rice yield in the world occurs in the Murrumbidge region of Australia, where rice is grown only once in five years, the land being under legumes during the other four years. After the discovery of chemical fertilizers in 1850, the yield of wheat in the UK went up to 28 bushels per acre. The beginning of this century saw the rediscovery of Mendel's laws of inheritance and the birth of the science of genetics and this provided new tools for tailoring plants to the needs of intensive agriculture.

An unique asset of tropical and sub-tropical agriculture is abundant sunlight all through the year, a factor which enables the cultivation of two or more crops in a year in lands with irrigation facilities. We in India, therefore, should aim at increasing productivity per hactare per day by growing green plants for as long a period as possible in a year. For achieving this aim, a radical reconstruction of the morphology and physiology of crop plants is necessary. During the last few years, varieties of crop plants possessing the type of morphological architecture and developmental rythm favourable for the efficient utilization of sunlight, water and fertilizer and consequently with a high yield potential, have been evolved through the use of genetic tools. I shall make a brief reference to a few such varieties, which have opened up new vistas in crop yields.

I. Rice : Rice occupies nearly 35 million hectares in our country but the average yield is only 1·1 tonnes per hectare, in contrast to over 4 tonnes per hectare in Japan. The varieties cultivated by us mostly belong to sub-species *indica* or *Oryza sativa*. Some of the reasons for the low yield potential of our rice culture are : (*a*) the week and tall straw of the *indica* varieties which makes the cultivation of rice under good conditions of soil fertility and the application of fertilizer difficult without causing lodging, (*b*) poor photo-synthesis duet o the extensive cultivation of rice during the monsoon when the sky is cloudy

---

*Indian Agricultural Research Institute, New Delhi.

during most parts of the day, (*c*) poor utilization of sunlight due to the shading of lower leaves by the upper ones, (*d*) poor soil and water management, and (*e*) lack of attention to details in cultural practices such as spacing and depth of transplanting. Some years ago, Chinese Scientists discovered a spontaneous mutant in the variety Dee-gee-Woo-gen which had the following characteristics :

(*i*) a dwarf plant habit, the plant attaining a height of about 60 cms,

(*ii*) stiff and erect leaves, enabling the maximum interception of sunlight;

(*iii*) insensitivity to photoperiod enabling the cultivation of the crop in any season;

(*iv*) absence of seed dormancy, rendering sowing possible immediately after harvest. Using this mutant, scientists in several parts of the tropics have now developed fertilizer-responsive and photo-insensitive varieties, which have revealed enormous possibilities for increasing the yields of *indica* rices.

Taichung Native 1, developed in Taiwan, is the first outstanding dwarf variety developed in this way in *indica* rice. This variety helped to dispel the old notation that only *japonica* varieties of rice are capable of responding well to the application of fertilizers. T.N. 1 was developed by crossing a tall *indica*, Tsai Yuen-Chung with Dee-gee-Woo-gen. IR-8, a variety developed at the International Rice Research Institute, The Philippines, from a cross between Peta, a tall *indica* variety and Dee-gee-Woo-gen, has also shown wide adaptability and has yielded 8 to 10 tonnes per hectare in many countries in South East Asia including India.

Several tropical *japonica* varieties developed in Taiwan such as Tainan 3, *Taichung* 65 and *Kaohsiung* 68 have done well in India and have yielded 5 to 7 tonnes per hectare. They are very resistant to the bacterial blight disease but have unfortunately sticky grains, arising from a low amylose content. Crosses between *japonica* and *indica* varieties have been attempted during the last two decades but due to difficulties in getting a wide spectrum of recombination, it has not been possible to develop many varieties with the desirable characters of both the parents. Also, it is now apparent that morphological more than physiological factors limit the response of plants to fertilizer application. There are, however, a few outstanding varieties which have resulted from the *japonica-indica* hybridization programme. ADT-27 developed in Madras State from the cross Norin 8 (*japonica*) × GEB-24 (*indica*) has yielded on an average 5 tonnes per hectare in the Thanjavur district of Madras State. This strain is now replacing the popular Kuruvai strain ADT-20. Sown in June or July, ADT-27 comes to harvest in 105 days after sowing. The grains are medium fine with white rice of good cooking quality. This variety has also enabled the introduction of double cropping in large areas in the Thanjavur district. Crosses between tropical *japonicas* and *indicas* may be more successful than those involving the temperate *japonicas* and need to be attempted on a large scale. Thus, it is now possible to get 8 to 10 tonnes of rice per hectare per crop in contrast to the 4 to 5 tonnes per hectare regarded possible earlier. What is more important, the barrier imposed by photosensitivity on repatterning rice growing seasons has been broken and sowing and harvesting periods can now be altered so as to synchronise crop growth with low cloudiness and abundant sunshine. Professor S.M. Sarkar has recently brought out clearly the importance of reconsidering the traditional rice sowing seasons and altering them to suit the needs of physiological efficiency. Given adequate attention, rice is one crop where India can increase production very substantially at once. About 14 million hectares of rice land have assured water supply and it should be possible to produce at least an additional 10 million tonnes of rice from this area alone during 1968, solely by increasing the efficiency of farming.

2. **Wheat** India produces only 10 million tonnes of wheat from about 13 million hectares. The reasons for the low yield even in lands which can be irrigated are : (*a*) the tall plant habit which results in lodging in fertile soils and (*b*) the rapid rise in temperature in February-March which imposes a ceiling on grain development and creates problems of soil and atmospheric drought. The second factor is so important that Sir Albert Howard remarked five decades ago that "Wheat cultivation in India is a gamble in temperature". The discovery in Japan of genes in the Norin wheat variety, which confer a dwarf and non-lodging plant habit, hence opened the door to the reconstruction of the morphology of the wheat plant. Several dwarfing genes were known for a long time in wheat, such as the S or C loci which govern the *Sphaerococcum* and *Compactum* characteristics respectively. These loci, however, have a pleiotropic effect on the ear, making it very dense and compact. The first variety which appeared to have the desired combination of short plant height, lodging resistance and kernel type was *Norin* 10. This variety was one of a collection of Japanese wheats brought to the United States by Dr. S.C. Salmon in 1948. Three recessive genes for dwarfing, with additive effect, have so far been identified.

Using the Norin dwarfing genes, the dwarf winter wheat variety, Gaines, was developed by Dr. C.A. Vogel in the Washington State, United States. Similarly, dwarf spring wheat varieties had been developed in Mexico by Dr. N.E. Borlaug and co-workers. In order to develop dwarf wheat varieties suitable for cultivation in India, the Indian Agricultural Research Institute, New Delhi, introduced in 1963 a large variety of wheat material containing the Norin dwarfing genes from Mexico through the courtesy of the Rockefeller Foundation and the Mexican Ministry of Agriculture. Even in the first year, i.e. 1963, the I.A.R.I. ditributed this material to several wheat breeding centres in order to assess the extent of their adaptation and amber grains, profuse tillering and a very high yield potential. It is derived from the cross, Penjamosils × Gabo 55. The Chapati and bread making properties are good and this strain is likely to become one of the most widely grown in the country. From the same cross, a variety named Siete Cerros 66 has been released in Mexico and a strain named Mexipak 65 in Pakistan. In Mexico, a sister selection with red grains is being grown commercially under the name SuperX. The same sister selection is called Indus 66 in Pakistan and V. 18 and P.V. 18 in India. Thus, this cross has yielded many outstanding selections characterised by a high yield potential and wide adaptability.

Sonalika: This is a single gene dwarf derived from the Mexican cross (II53-388-An) (Yt. 54 × N 10 B) Lr. III 8427 and the original material was received under the number S. 308. The grains of this variety are bold and amber. It possesses resistance to all the three rusts and does well both under timely-sown and late-sown conditions. This selection was made at the I.A.R.I.

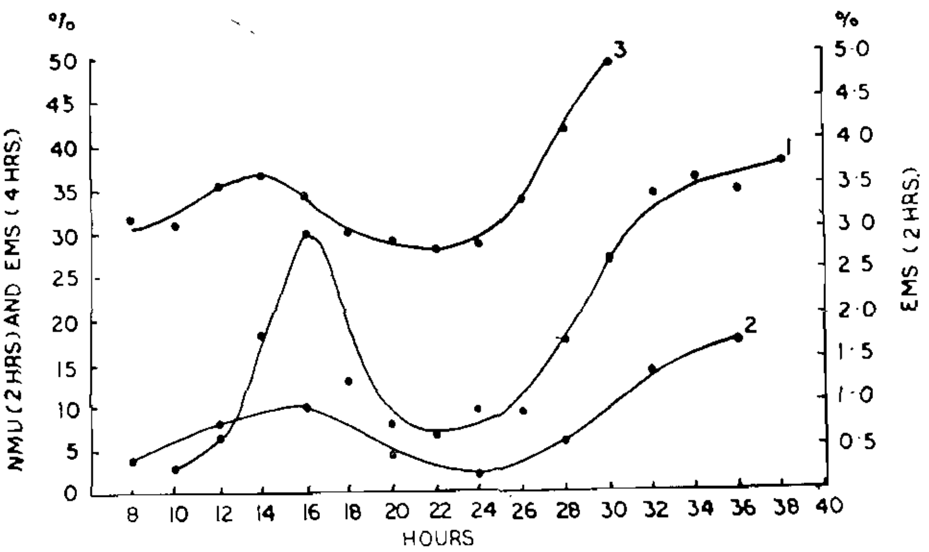
Safed Lerma: This is also a single gene dwarf derived from the Maxican cross (Y 50 × N.10B) *L.52LR3*. Since it is from material backcrossed thrice to *Lerma Rojo* 64A, it resembles *Lerma Rojo* 64A very closely in height, pigmentation and other characters. The main difference lies in the white semi-hard nature of this strain, in contrast to the red and soft grains of *Lerma Rojo* 64A.

Chhoti Lerma: This is a white seeded, two-gene dwarf derived from the Mexican cross (R 64 sib × HUA.R.) It is highly resistant to lodging and has a high degree of resistance to all the three rusts.

The age of Algeny: The late Prof. H.J. Muller initiated in 1927 what has been aptly termed recently by Professor J. Lederberg as the era of "Algeny", a term coined to indicate the artificial transmutation of the gene or genetic alchemy. During the last 40 years, much knowledge has been gathered on techniques useful for increasing the frequency as well as for widening the spectrum of induced mutations. With the advent of molecular genetics, a wide range of powerful chemical mutagens have also become available. Alkylating agents like ethylene imine, diethyl sulphate and ethyl methane sulphonate have proved to be very efficient mutagens in a wide range of plants. Chemicals like Nitrose-methyl urea and Nitroso-ethyl-urea have proved to be even more potent and have been described by Prof. A. Rapaport of the USSR as "super-mutagens".

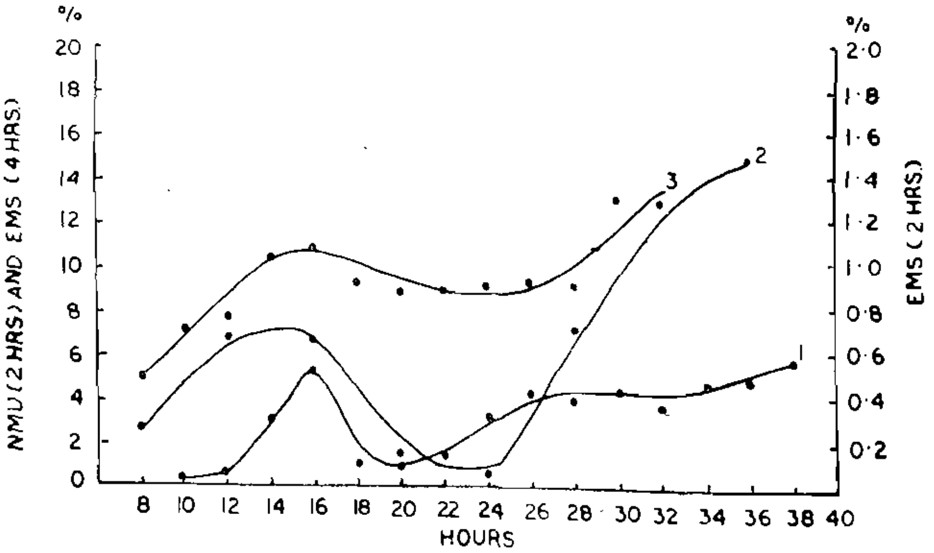
The numerous biological, environmental and pre-and post-treatment parameters which influence the efficiency of mutagenic treatments are also now fairly well understood. The mutation frequency can be enhanced several folds by altering factors such as moisture content of seeds, oxygen availability and treatment temperature. In the case of chemical mutagens like EMS and

SPIKE PROGENIES



1 - EMS 2 HRS.
2 - EMS 4 HRS.
3 - NMU 2 HRS.

FREQUENCY OF MUTATION



NMU, the effect is greatly increased if the treatment period coincides with the "S" phase of DNA synthesis. This synchronisation can be brought about by carrying out the treatments of seeds pre-soaked in water for specific periods. For example, the S-phase commences in barley seeds after 16 hours of soaking in water at 25°C. This can be varified through studies relating to the incorporation of $3H$-thymidine. When the mutagen treatment is started after pre-soaking seeds in water for 14 to 16 hours, the mutation frequency is considerably increased (Fig. 1).

Though the frequency of mutations can be manipulated in different ways, the spectrum of mutation is by and large random. It has not so far been possible to obtain any great degree of control over the mutation spectrum in higher plants. Some experiments currently underway at the Indian Agricultural Research Institute may help in achieving a greater degree of specificity in the induced mutation spectrum. The technique devised by me for this purpose is based on the principle that DNA synthesis along a chromosome is not synchronous, thereby affording an opportunity for affecting different parts of the chromosome differently through short duration treatments. These experiments represent the first major advance in gaining a fairly precise control over the type of mutation induced, the earlier experiments in this field having largely uttempted to alter the spectrum through a change in the mutagen used.

I shall cite just two examples from the recent work carried out at the I.A.R.I. to indicate what an invaluable supplement mutation breeding has now become in crop impovement through the breeding of new strains.

Sharbati Sonora: The 2-gene dwarf wheat variety Sonora 64 introduced by the I.A.R.I. in 1963 has, as mentioned earlier, proved to be capable of yielding about 6 tonnes per hectare when grown properly. The seeds of this variety are, however, red in colour and hence fetch a little lower price in the grain market. The variety was subjected to mutation breeding in 1963 by Dr. G. Varghese and myself and an amber seeded mutant isolated from γ-ray treated material was approved for release by the Central Variety Release Committee in 1967 under the name "Sharbati Sonora". Sharbati Sonora resembles Sonora 64 in all other respects except the quality and the colour of grains. Besides having bold and amber seeds, Sharbati Sonora was found to possess on an average 16·5% protein in contrast to about 14·0% in the parent strain. Also, Sharbati Sonora has about 4-5 grams lysine in 100 grams protein, in contrast to only about 2 grams lysine in 100 grams protein in Sonora 64. It appears that a major factor for protein and lysine synthesis may be located near the locus concerned with grain colour. This mutant has helped to disprove the idea that protein quality and quantity cannot be coincidentally improved. Since Sharbati Sonora yields over 6 tonnes per hectare, it is also a symbol of the possibility of developing high yielding-cum-high quality varieties of cereals and other crops.

reactions to the races of black, brown and yellow rusts prevalent in the country. In addition to breeding material, bulk quantities of four commercial Spring Wheat varieties—*Lerma Rojo* 64A, *Sonora* 63, *Sonora* 64 and Mayo 64 were also obtained. These varieties have been tested in all the wheat growing States of India during the last four *rabi* seasons under the All India Co-ordinated Wheat Improvement Project. In addition, they have been subjected to detailed physiological, pathological, chemical and agronomic tests at the Indian Agricultural Research Institute. Two of these varieties, *Lerma Rojo* 64A and *Sonora* 64 were approved by the Central variety Release Committee of the Indian Council of Agricultural Research in 1965 for cultivation in irrigated areas.

Lerma Rojo is a late variety with a high degree of resistance to yellow rust. It performs very well under timely sown conditions and in areas characterised by yellow rust epidemics. *Sonora* 64, on the other hand, is an early variety and is well suited for being grown in rotations like maize-wheat, potato-wheat, rice-wheat, sugarcane-wheat etc. (Tables 1 & 2). It has two genes for dwarfing and hence it is the most lodging-resistant variety so far released. Being early, it is a safe variety for cultivation under high fertility conditions in the eastern parts of U.P., Bihar, West Bengal, Rajasthan, Madhya Pradesh, Gujarat, Maharashtra, Orissa and Madras. *Sonora* 64 should not be sown before the middle of November in areas where the normal sowing time is late October or early November. It is susceptible to yellow rust and hence is not recommended for areas where yellow rust appears in an epidemic conditions.

New dwarf wheats : Varietal diversity as well as a rapid replacement of varieties are essential for sustaining high wheat yield over many seasons. From the advanced generation material received from Mexico in 1963, several selections such as S. 227, S. 307 and S. 308 have been found to perform very well in the Northern Plains zone. These strains have amber grains and a very high yield potential. The original material of S. 227 received from Mexico segregated for resistance to brown rust and resistant selections were made. The highest yield in national demonstrations in 1965-66 and 1966-67 were obtained with S. 227, which yielded respectively 68 and 82 quintals per hectare in a farmer's field in the Delhi State. S. 307 is a derivative of a cross involving *Lerma Rojo* and the Japanese dwarf strain *Norin* 10-B. Seeds of S. 227, S. 307 and S. 308 were multiplied at Wellington, in the Nilgiri Hills during the summer of 1966 and 1967 and these varieties were approved for general cultivation in 1967. A brief description of these new strains and the names given to them are given below :

Kalyan Sona : This is a selection possessing resistance to brown rust made from the population of S. 227 grown in 1963-64. Selections were made at the I.A.R.I., the Punjab Agricultural University, Ludhiana and the U.P. Agricultural University, Pantnagar. This is a strain with medium maturity,

Conversion of Japonica rice into Indica type: As mentioned already, some of tropical *japonica* variet.es of rice bred in Taiwan such as Taichung 65, Tainan 3 and Kaoschung 68 are doing well in several parts of India. These strains are also highly resistant to bacterial blight. They will hence become popular provided the sticky characteristics of the grains are removed. Most *japonica* varieties have a low amylose content and a low gelatinization temperature. Since desirable recombinations have been rare in *japonica* × *indica* hybrids, the variety Taichung 65 was subjected to mutation breeding by Dr. E.A. Siddiq and myself in 1965. Several mutants have been isolated which possess all the plant characters of Taichung 65 and the grain characters of *indica* strains. Some of the *indica* mutants also have about 14% protein content. These are being multiplied and tested for yielding ability. This technique therefore has opened a way for rendering *japonica* varieties acceptable to Indian farmers and consumers. Further, the induced *indica* mutants in *japonica* varieties and *japonica* mutants in *indica* varieties have helped to understand the mode of sub-specific differentiation in *Oryza sativa*. The data suggest that the differentiation has proceeded largely through the accumulation of independent mutants into coherence groups, probably under the influence of disruptive selection. This has resulted in the emergence of a whole constellation of characters, commonly found to be associated with the *japonica* group.

The age of algeny has thus opened up altogether new possibilities in the development of high yielding varieties and in the genetic manipulation of protein quantity and amino acid balance. In some fields of plant breeding like the exploitation of heterosis, a theoretical understanding of the phenomenon has not been essential for achieving significant practical gains. In mutation breeding, on the other hand, practical accomplishments will grow with a growth in our understanding of the mechanisms underlying the mutation process.

Genetic manipulation of quality: The development of varieties with a high genetic potential for yield and an agronomy capable of bringing out their potential have rendered the prospects of achieving quantitative self-sufficiency in food extremely bright. Serious consideration should hence be given hereafter to the improvement of the quality of food, particularly with a view to eliminate the widespread protein under-nourishment. The possibility of tackling the problem successfully in the immediate future through increased production of milk, eggs and other animal products is remote, since the efficiency of conversion of plant foods into edible products by the animal is extremely low, seldom exceeding 40%. Even now the total proportion of cows in the bovine population is low in most parts of India. The plant—animal—man food chain widely in vogue in the developed nations would be difficult to achieve in the near future in our country, owing to the shortage of grains for even the human population. The more practical approach would be to improve the protein quality of cereals, pulses and millets through genetic engineering, so that the consumer can get better quality food without any extra expenditure. That this can be done was demonstrated elegantly by Prof. E.T. Metz and his colleagues at the Purdue University in the United States in 1964, who found a high lysine content to be associated with the opaque-2 and floury-2 genotypes in maize. They further found that in rats fed with a diet containing 90% opaque-2 maize for 28 days the average gain *in weight was 97 grams while in control rats fed on a standard hybrid maize*, the average gain was 27 grams.

Recent research at the Indian Agricultural Research Institute by Dr. A. Austin and his colleagues has revealed enormous possibilities for increasing the quantity of proteins in rice and wheat. Strains of rice with 14 to 15% protein content have been isolated in material treated with mutagens. With the generous help of the Rockefeller Foundation, an Amino Acid Analysis Laboratory has now been set up which has made rapid progress in the development of cereals and pulses high in the content of some essential amino acids such as lysine, methionine and theonine possible. Thus, the wheat variety Sharbati Sonora produced by the technique of algeny has been found to have twice the lysine content of the parent variety, Sonora 64. This mutant has further shown that contrary to some earlier opinion, protein quantity and lysine content are not negatively associated. Also, it is hoped that through cooperative experiments with the Nutrition Research Laboratory, Hyderabad, certain defects in the grains of *Jowar* and *Khesari dal* (*Lathyrus sativus*) which are responsible for causing the diseases pellagra and lathyrism respectively can be rectified soon. Concerted attempts are needed to change the current market classification of quality of cereals based on characters such as colour and lustre. The price of wheat, rice, *jowar*, maize and *bajra* should instead be fixed on the basis of the quantity and quality of proteins present. The introduction of a protein premium in the pricing policy would be the most effective way of arousing a protein consiousness in the rural areas. The results already obtained encourage the hope that the High-Yielding-Varieties Programme would soon become a High-Yielding-cum-High-Quality-Varieties Programme.

These examples would be sufficient to indicate the pace of progress of genetic research and the vast panorama of exciting possibilities it has opened up in Indian Agriculture. By growing two crops—either the dwarf varieties of wheat or rice and the hybrids of maize, Sorghum or Pearl Millet —in the same field along with suitable agronomic practices, nearly 10 tonnes of grains can be produced per hectare per year. This yield can go up to 20 tonnes per hectare, if 3 or 4 crops are raised through the adoption of suitable multiple or relay cropping techniques. India has nearly 100 million hectares under food crops and thus production potential of our agriculture and the scope for evolving new land use patterns are vast.

Dr. V.G. Panse has worked with distinction and single minded devotion throughout his life for the application of genetics and biometrics in crop improvement. The recent results offer eloquent testimony to his foresight and vision.

TABLE 1

Productivity per day in some wheat varieties

Variety	*Days in the field*	*Productivity (Kg/ha/day)*
NP 880	151	17·1
Lerma Rojo	153	24·8
S·227	154	34·2
Sonora 64	133	42·5

TABLE 2

Wheat yield in rotations involving late sowing

Location	*Rotation*	*Wheat variety*	*Date of sowing*	*Yield Q/ha.*
Delhi	Jowar-Wheat	Sonora 64	29.12.64	42·4
Jullundur	Potato-Wheat	Sonora 64	2. 1.66	40·9
		Lerma Rojo	2. 1.66	36·2
		Sonora 64	15. 1.66	34·4
		Lerma Rojo	15. 1.66	33·6
Samastipur	Rice-Wheat	Sonora 64	5.12.65	37·1
Darbhanga	Rice-Wheat	Sonora 64	19.12.65	25·5
Shahabad	Rice-Wheat	Sonora 64	29.12.65	23·2
Delhi	Paddy-Wheat	Sonora 64	4.12.65	55·3

Bibliography

List of Scientific Papers Contributed by Dr. V. G. Panse

1933

1. (**and Wad, Y.D.**) : The nitrogen balance in the black cotton soils on the Malwa plateau, Ind. J. Agric. Sci., Vol. 3, 820.

1934

2. (**with Jackson, F.K. and Wad, Y.D.**) : The supply of humus to soils, Emp. Cotton Grow. Rev. Vol. 11, 111.

1935

3. (**and Hutchinson, J.B.**) : Sampling for staple-length determination in cotton trials, Ind. J. Agric. Sci., Vol. 5, 545.

4. (**and Hutchinson, J.B.**) : Size, shape and arrangement of plots in cotton trials, Ind. J. Agric. Sci., Vol. 5, 524.

5. (**and Hutchinson, J.B.**) : An application of the method of covariance to selection for disease-resistance in cotton, Ind. J. Agric. Sci., Vol. 5, 554.

6. (**and Hutchinson, J.B.**) : A study of margin effect in variety trials with cotton and wheat, Ind. J. Agric. Sci., Vol. 5, 671.

1936

7. (**and Hutchinson, J.B.**) : The introduction of improved strains of crop plants in Central India and Rajputana, Agric. and Livestock Ind. Vol. 6, 397.

1937

8. The delimitation of areas for strains of agricultural crops with special reference to cotton, Proc. Ist Conference of Scientific Research Workers on Cotton in India, 411.

9. (**and Hutchinson, J.B.**) : An examination of an analysis of serial experiment, Agric. and Livestock Ind., Vol. 7, 332.

10. (**and Hutchinson, J.B.**) : *On an attempt to use* hand-spinning for testing quality in cotton, Agric. and Livestock Ind., Vol. 7, 339.

11. (**and Patel, A.F.**) : A *genetical study* of roots in relation to disease-resistance in cotton, Ind. J. Agric. Sci., Vol. 7. 451.

12. (**and Hutchinson, J.B.**) : The design of field tests of plant breeding *material*, Ind. *J.* Agric. Sci., Vol. 7, 531.

1938

13. Preliminary studies on sampling in field experiments, Proc. 1st Session of the Indian Statistical Conference, Cal., 139.

14. (**with Hutchinson, J.B. and others**) : Crop *analysis and varietal* improvement in Malvi Jowar, Indian J. Agric. Sci., Vol. 8, 131.

15. Methods of staple-length *determination in cotton* used at different plant breeding *stations in India*, Ind. J. Agric. Sci., Vol. 8, 582.

16. (**with Hutchinson, J.B. and G.K. Govande**) : The inheritance of *agricultural characters in* three inter-strain crosses in cotton, Ind. J. Agric. Sci., Vol. 8, 757.

1939

17. (**and Ramiah, K.**) : A reply to Dr. Mason's note on plant breeding technique, Emp. Cott. Grow. Rev., Vol. 16, 25.

1940

18. Inheritance of quantitative characters and plant breeding, J. Genet. Vol. 40, 283.

19. A statistical study of *quantitative inheritance*, Annals of Eugenics, Vol. 10, 76.

1941

20. A statistical study of the relation between quality and return per acre in cotton, Ind. J. Agric. Sci., Vol. 11, 546.

21. Size and shape of blocks and arrangements of plots *in* cotton trials, Ind. J. Agric. Sci., Vol. 11, 850.

1942

22. Methods in plant breeding, Ind. J. *Genetics and* Plant Breeding, Vol. 2, 151.

23. (**and Mokashi, V.K.**) : Manuring of rainfed Cotton in India, Indian Cotton Grow. Rev., Vol. 6, 1-9.

1943

24. (**and Sahasrabudhe, V. B.**) : A rapid method of sampling for fibreweight determination, Ind. J. Genetics and Plant Breeding, Vol. 3, 28.
25. (**and Sukhatme, P. V.**): Size of experiments for testing seed or vaccines, Ind. J. Vet. Sci. and Animal Husb., Vol. 13, 75.

1944

26. (**and Kalamkar, R. J.**): Forecasting and estimation of crop yields, Current Science, Vol. 13, 120.
27. Manuring of cotton in India, Indian Farming, Vol.5, 131.
28. (**and Kalamkar, R. J.**): A further note on estimation of crop yields, Current Science, Vol. 13, 223.
29. (**with Nanda, D. N. and Mohammad Afzal**): A statistical study of flower production in cotton, Ind. J. Agric. Sci., Vol. 14,78.
30. (**and Ayachit, G.R.**): Tables of 10 per cent point of Z and the variance ratio, Ind. J. Agric. Sci , Vol. 14, 244.

1945

31. (**with Kalamkar, R. J. and Shaligram, G. C.**) A large scale yield survey on cotton, Current Science, Vol. 14, 287.

1946

32. Manuring of cotton in India—A review of the results of cotton manurial trials carried out in India with suggestions for future experiments, *Indian Central Cotton Committee, Bombay.*
33. An application of the discriminant function for selection in poultry. J. Genetics, Vol. 47, 212.
34. Report on the scheme for the improvement of Agricultural Statistics, Imperial Council of Agricultural Research, New Delhi.
35. Plot size in yield surveys on Cotton, Current Science, Vol. 15, 218.

1947

36. (**and Sahasrabudhe, V. B.**) : Yield of rainfed cotton and its improvement, Ind. Cott. Grow. Rev., Vol. 1, 10.

37. Plot size in yield surveys, Nature, Vol. 159, No. 4050, 820.

38. (with **Dandawate, M. D.** and **Bokil, S. D.**): Summary of past experimental work on wheat, millets, oilseeds and pulses, Appendix II (*a*). Report on soil fertility investigations in India, by A. B. Stewart, Imperial Council of Agricultural Research, New Delhi.

39. (and **Shaligram, G. C.**): Improvement of Statistics of Cotton Production in India, Ind. Cott. Grow. Rev., Vol. 1, 199.

1948

40. (and **Khargonkar, S.A.**): Effect of seed quality in yield of cotton, Ind. Cott. Grow. Rev., Vol. 2. 17.

41. (and **Sukhatme, P. V.**): Crop Surveys in India—I, J. Ind. Soc. Agric. Stat,, Vol. 1, 34.

42. (and **Bokil, S. D.**): Estimation of genetic variability in plants, J. Ind. Soc. Agric. Stat., Vol. 1, 80.

1949

43. Cotton in Egypt, Ind. Cott. Grow. Rev., Vol. 3, 1.

44. (and **Khargonkar, S. A.**): Discriminant function for selection of yield in cotton, Ind. Cott. Grow. Rev., Vol. 3, No. 4, 179-186.

1950

45. (with **Sahasrabudhe, V. B.** and **Mokashi, V. K.**): The new series of manurial trials on rainfed cotton in India, Ind. Cott. Grow. Rev., Vol. 4, 95-105.

1951

46. (with **Shaligram, G.C.** and **Bokil, S.D.**): Improvements of yield forecasts of cotton, Ind. Cott. Grow. Rev., Vol. 5, 181-201.

47. (with **Sahasrabudhe, V. B.** and **Mokashi, V. K**): Coordinated manurial trials of rainfed cotton in Peninsular India, Ind. Jour. of Agric. Sci., Vol. 21, 113-135.

48. (and **Sukhatme, P.V.**): Crop surveys in India—II, Jour. Ind. Soc. Agric. Stat., Vol., 3, 97-163.

49. (and **Koshal, R. S.**): A field investigation of the causes of reduction of cotton area in Bombay State, Ind. Cott. Grow. Rev., Vol., 5, 61-99.

50. **(and Rao, V.R.)** : System of Agricultural Statistics in India, International Statistical Conference, India.

1952

51. Trends in areas and yields of principal crops in India, Agricultural Situation in India, Vol. 7, 144-148.

1953

52. Trends in land utilization in India (with special reference to fallow land), Agricultural Situation in India, Vol. 8, 21-44.

53. **(with Yates, F. and Finney, D.J.)** : The use of fertilizers on food grains, I.C.A.R. Research Series No. 1.

54. Designs for plant breeding agronomic experiments in rice, Meeting of the Working Party in Fertilizers, International Rice Commission, Bankok.

55. **(and Sukhatme, P.V.)** : Experiments on cultivator's fields, J. Ind. Soc. Agric. Stat., Vol. 5, 144.

1954

56. Estimation of crop yields, Bulletin of the Food and Agricultural Organisation of the United Nations, Rome.

1955

57. Estimation of cost of production of crops, Published by the Indian Central Cotton Committee, Bombay.

58. **(and Sukhatme, P.V.)** : Statistical methods for agricultural workers, Published by the Indian Council of Agricultural Research, New Delhi.

59. Some Statistical Problems, Presidential Address at the 42nd Indian Science Congress, Baroda Session.

1956

60. A pilot survey for the estimation of the cost of milk production under Indian Conditions, Vol. 1, XIVth International Dairy Congress Proceedings, Rome, 1-26.

1957

61. Genetics of quantitative characters in relation to plant breeding, Ind. J. of Genetics and Plant Breeding, Vol. 17, No. 2, 318-328.

62. Some observations on the problems of estimating farm costs, Proceedings of the FAO Development Centre on Farm Planning and Management for Asia and the Far East, New Delhi.

63. Some comments on the objectives and method of 1960 World Census of agriculture, International Statistical Institute session at Stockholm.

1958

64. **(with Sukatme P.V. and Sastry, K.V.R.):** Sampling technique for estimating the catch of sea fish in India, Biometrics, Vol. 14, No. 1, 78-96.

65. Status of agricultural experiments, Food and Agricultural Organisation of United Nations, Rome.

66. Problems and Techniques in the study of the cost of production in agriculture, Ind. J. Agri. Eco., Vol. 13, No. 3, 1-12.

1959

67. Recent trends in the yield of rice and wheat in India, Ind. J. Agric. Eco., Vol. 14, No. 1.

1960

68. **(and Abraham, T.P.)** : Simple scientific experiments on farmer's land, International Statistical Institute, 31st Session at Brussels, 3-13.

69. Regional Planning and Agricultural Census, Manila Seminar.

70. **(with Amble, V.N. and Puri, T.R.)** : Statistical control of operational efficiency in rinderpest eradication compaign, Sankhya, Series B, Vol. 23, 1-12.

71. **(and Bokil, S.D.)** : Design of cost studies in agriculture, Seminar on Problems of Cost Studies in Agriculture at Matheran, Oct.-Nov. 1960.

72. Sample survey for quality control of cotton, Ind. Cott. Grow. Rev. Vol. 14, No. 4, 317-321.

73. National Sample Survey, Agricultural Statistics and Planning in India, Changing India, Published by Gokhale Institute of Politics and Economics.

74. Some aspects of the current fertilizers situation in Japan, Agricultural Situation in India.

75. (**and Menon, V.S.**) : Index numbers of Agricultural Production, in India, Ind. J. Agric. Eco., Vol. 16, No. 2, 18-36.

76. Present and Future of Agricultural Planning in India, A.I.C.C. Economic Rev., 1-6.

77. Food Availability and requirement for India, International Statistical Institute, 33rd Session, Paris, 1-10.

78. (**with Amble, V.N. and Puri, T.R.**) : Cost of Milk Production in Delhi 1953-55, I.C.A.R. Bull.

1962

79. Prospects for India's Agriculture, A.I.C.C. Economic Rev., 1-7.

80. World Agricultural Census, Agricultural Situation in India.

81. (**and Khanna, R.C.**) : Manuring of Cotton in India (A review of the results of cotton manuring trials carried out in India with suggestions for future experiments).

1963

82. Thirty Years of Statistics in Agriculture in India, I.C.A.R. Rev. Ser. No. 27.

83. (**with Amble, V.N. and Raut, K.C.**) : Cost of Milk Production in Madras State 1957-59, I.C.A.R. Bull.

84. Comparison of yield estimate prepared on the basis of traditional and crop-cutting methods—A rejoinder, Ind. J. Agric. Econ. Vol., 18, No. 2, 33-36.

85. (**and Khanna, R.C.**) : Response to some important Indian crops to fertilizers and factors influencing this response, Paper read at ECAFE Conference for setting up fertilizer industry in Asia and the Far East, Bombay, Nov. Dec.-1963.

86. Plot size again, J. Ind. Soc. Agric. Stat., Vol. 15, 151-159.

87. Yield trends of rice and wheat in first two Five Year-Plans in India, J. Ind. Soc. Agric. Stat., Vol. 16, No. 1, 1-50.

88. (**With Amble, V.N. and Raut, K.C.**) : Cost of production and price of milk, Gosamvardhan.

89. (**with Abraham, T.P. and Leelawathi, C.R.**) : Yardsticks of additional production of certain foodgrains, commercial and oil-seed crops, I.C.A.R. Bull.

90. (**with Amble, V.N. and Abraham, T.P.**) : A plan for improvement of nutrition of India's population, Ind. J. Agric. Econ., Vol. 19, No. 8.

91. (**with Singh, D. and Murty, V.V.R.**) : Sample Survey for estimation of milk production (Punjab), I.C.A.R. Bull.

92. Sample survey for estimation of area and yield of coconut and arecanut, 1959-60 to 1961-62, Arecanut Journal, 161-172.

1965

93. (**and Amble, V.N.**) : The future of India's population and food supply, United Nations World Population Conference, Belgrade, Yugoslavia, 30 August to 10 Sept., 1965.

94. (**with Amble, V.N. and Raut, K.C.**) : The cost of milk to the producer and the consumer, Indian Livestock, 37-39 and 47-48.

95. Fertilizer Trials, Fertilizer News, Vol. 10, No 12.

96. Fertilizer Recommendations, National Seminar on Fertilizers.

97. Some problems of Agricultural Census taking with special reference to developing countries, FAO, Rome.

1966

98. (**and Singh, D.**) : Promotion and assessment of Technological change in Indian Agriculture, Ind. J. Agric. Econ. Vol. 21, No. 1.

99. (**with Rajagopalan, M. and Pillai, S.S.**) : Estimation of Crop yields for small areas, Biometrics, Vol. 22, No. 2 (July).

100. (**with Singh, D. and Murty, V.V.R.**) : Sample survey for estimation of milk production (U.P.) I.C.A.R. Bull.

101. (**with Amble, V.N. and Raut, K.C.**): Cost of milk production in West Bengal, I.C.A.R. Report Series No. 28.

102. (**with Murty, V.V.R. and Sathe, K.V.**) : Aerial photography for determining land use—Jour. Ind. Soc. Agric. Stat. Vol. XIX, Dec. 1967, No. 2, 25-35.