A STUDY OF STATISTICAL METHODOLOGY IN FARM COST AND MANAGEMENT SURVEYS

BY

SUDARSHAN KUMAR RAHEJA

Prepared under the guidance of Dr. V.G. Panse, Statistical Adviser, Indian Council of Agricultural Research.

Dissertation submitted in part fulfilment for the award of Diploma in Agricultural Statistics, Indian Council of Agricultural Research

INDIAN COUNCIL OF AGRICULTURAL RESEARCH,

NEW DELHI

1957

C O N T E N T S

1. INTRODUCTION

1.1 Farm Management Defined

Various agricultural economists have defined farm management in various ways. According to one definition, 'It is a science dealing with the combination and operation of production factors including land, labour and capital, and selection of the kinds and amounts of crops and livestock enterprises which will provide maximum and continuous returns to the farm unit'. A second definition is 'Finding good solutions to the economic questions of production that a farmer faces everyday on the land'. There can be other definitions depending upon how we look at the subject. But brushing aside the question of defining it technically, it may be very well stated to be, 'Finding sound means for improving the farmers' lot'. The agricultural situation in India is rather poor and has been badly neglected until very recently. This can be seen from the following account of Indian Agriculture :

'Of the total (agricultural) population of about 250 millions, 71 millions or about 29% are self supporting persons, 147 millions or roughly 59% are non-earning dependents and the remaining about 12% are earning dependents. Of the self supporting 71 millions, owner cultivators are roughly 45.7 millions, tenant cultivators 8.8 millions, cultivating labourers 14.9 millions and agricultural rentiers 1.6 millions. It is the high percentage of non-earning dependents in the agricultural sector that is responsible, to a large extent, for the backwardness and poverty of rural masses'.

Land in India constitutes by far the largest portion of the natural resources of the country. Similarly agricultural produce constitutes the largest portion-nearly 50%-of national product. Also 83% of the Indian population lives in rural area, about 70% of whom are following agricultural porsuits i.e. 250 million Indians depend for their livelihood on agriculture alone. It is, therefore, of vital importance that development of agricultural practices on modern scientific lines be given first and foremost consideration. In fact, no economic and social progress of the country can be made without strengthening the agricultural base of our economy. We can thus realise the pressing need of organising a study to probe into the conditions under which farmers work and the problems they have to face from day to day. Agricultural operations in India are carried out primarily with the help of human and bullock labour and the implements used are little better than what they were centuries ago. Thus a mere recommendation of better farming practices with modern implements will not do. There is, in fact, a need for re-orientation of the entire agricultural system.

A suggestion that is bound to prove of immediate consequences is the setting up of educational institutions for imparting relevant agricultural knowledge to the farmers. Western countries are considered to be advanced in agriculture. The one major factor that has contributed significantly towards this advance is that the farmers in those countries are educated and can well appreciate the economic principles connected with agriculture. The only Asiatic country that is considered agriculturally advanced is Japan and even in that country, every farmer spends 7-8 years in getting himself educated and acquainted with the progress made by his counterpart in the other countries. According to Lewis (1957) 'A continous improvement in the managerial ability of farmers is necessary for the economic advancement of any country. In the long run, improved managerial ability of farmers will depend very heavily on more and better schooling for rural children. It will also depend upon more research, teaching and extension on the subject of farm management economics. One should bear in mind that there is a growing body of knowledge of the economic principles governing the organisation and operations of farms. One has, therefore, to do original and imaginative thinking on the application of principles of farm management to the farms of his region'.

From this we realise the need for providing adequate facilities for schooling for the farmers. This will go a long way in helping the farmer understand the basic principles and will make the task of agricultural economist much simpler.

1.2 Development of Farm Management Studies in Other Countries

1.2.1 U.S.A. Although some cost statements were available in the U.S.A. as early as 1870, seemingly due, primarily to the tendency of showing profits and not of emphasising the need of abandoning the old, out-of-date techniques, regular farm management studies and collection of data started in 1902 when Heys introduced the method known as the Route method, which later developed into the present Cost Accounting method, by which farmers were induced to keep day to day records of transactions on the farm with the help of a special agent. As this involved heavy costs on the part of the collecting agency as well as the farmer, other methods were tried. The notable among these was the Survey method introduced by G.F. Warren of Cornell University. This consists of use of an extensive questionnaire and a personal enquiry from the farmer by an investigator. As no book-keeping is required, the reliability of this method depends mainly upon whatever information the farmer can give from his memory.

1.2.2 U.K. In the U.K., regular farm management studies and collection of data on the Cost Accounting method started rather late in 1913 under the direction of C.S. Orwin at Wye. This method mainly has been in use in England since then. To obtain information on land and variation of farm incomes and costs of production, a continuous farm management survey based on farmers' accounts was initiated in England and Wales in 1936. The survey was connected with commercial undertakings where the farmer is wholly or mainly occupied with the work and management of his farm. The approach adopted by the Oxford Agricultural Economics Research Institute in England has, in general, the same pattern as that of the American Investigation into production costs. A notable point is that under this method, the nature of farm products consumed by the household is reckoned on the side of receipts as if they had been actually sold.

1.2.3 <u>Other European Countries</u> The publication of results of farm accounting by the Secretariate of Swiss Peasants, Switzerland, as early as 1901 proved useful in solving the technical problems. Laur (Switzerland) introduced an elaborate method of analytical accountancy. Accountancy offices on the Swiss model were established in other European countries like Sweden, Norway, Finland and Germany. The development of methodology both in collection as well as analysis of data was influenced more by the economic principles (The term Analytical Accountancy has been used in early stages instead of Cost Accounting). Thus, initially, only a method of accounting procedure was adopted as different from a regular system of Cost Accounting as known to-day. Many changes have occured in the procedure of farm costing in the U.S.A. In recent years, investigations have been mainly confined to the collection, analysis and presentation of quantitative cost data from which we can draw more dependable conclusions than the money cost data because of fewer changeable factors affecting the quantitative expenditure. The quantitative data is converted into values by the application of current prices or rates.

From the foregoing account it is clear that in the earlier years, investigations have mostly been cost studies only. The need for farm management investigations for bringing out the defects and limitations in the existing practices and helping the farmer raise his farm income by suggesting optimal allocation of farm resources was felt only recently.

Investigations have also been carried out by farmers, on private level, collectively. The College of Agriculture, Illinois, initiated in 1924, a new type of investigations on farm management called the farm bureau-farm management services, similar in nature to the co-operative farm of efficiency service which could also be used to provide data necessary for scientific research. A similar approach was adopted in Hampshire (England).

1.2.4 Japan From the above, it can be seen that in the advanced Western countries like the U.S.A. and U.K., farm management studies on scientific lines started in the beginning of the present century. In the East, the only country in which regular farm management studies, based on modern ideas, were made is Japan where, some ten years back, these studies were initiated with the help of Cost Accounting method. Although no methodology has yet been established there, the newly designed extension service system was inaugurated in 1948 and now there are many sided programmes being formulated and put into practice. In Japan, the way of thinking of extension of farm management techniques seems to be influenced by the objective of extending such techniques among farmers for maximising profits by proper combination of crops and livestock enterprises and well-arranged distribution of family and bullock labour and, therefore, more emphasis has been laid on book-keeping on the part of the farmer.

To popularise this, three steps have been taken :-

1. Simplified forms of book-keeping have been adopted to enable the farmer to understand and practise it with ease.

II. Greater attention is paid towards increasing the educational qualifications of the local government staff member in charge of extension services.

III. Extension of farm book-keeping practices has been done through farm advisers in the farm of group education e.g. book-keeping has been added to the projects of farm study groups.

The approach of propagating improved practices and techniques through modal farmers has also been adopted.

Although Cost Accounting method has been widely used, farm management analysis based on Survey method of investigations has helped Japan increase the productivity of resources used in agricultural production. It has in particular, helped bring about intensification of land use and land improvement and paved the way for the adoption of improved varieties of seeds and application of chemical fertilisers. It seems that the output conditions in Japan have reached a static point and little or no increase in the same is possible through the adoption of better farming practices. Replacement of bullocks by power tools is not being widely encouraged as it would mean requirements of less human labour which in turn would create unemployment. In fact the approach to the mechanisation is based on the assumption (already proved) that mechanisation would increase the demand for labour on the farm on a per acre basis though it would reduce the need for labour on a per-unit-of-crop basis. Farm management studies are, therefore, looked upon as a possible source for indicating some other means for increasing the output e.g. through the efficient managerial ability of farmers. Farm management studies are, therefore, given considerable importance.

1.2.5 India In India, no organised efforts seem to have been made in this direction until 1923. Whatever records existed were from land records of some of the states containing elaborate date on cost of cultivation based on rough method of estimation. Although a large number of rural surveys have been carried out in the past, the scope of such surveys was mostly confined to the study of economic problems of rural life and not of farm management. The ball was set rolling when in 1920, Indian Central Sugarcane Committee steressed the need for an enquiry in deciding in what areas the sugar industry had the best chances of development and obtaining accurate data on costs of various raw materials. As a result, an enquiry into the cost of production, spread over three crop years, was undertaken in 1933 in the principal sugarcane and cotton tracts in India. A general survey of economic position of the cultivator and general standdards of farm management still remained out of the scope of this enquiry in which cost accounting method was employed.

In 1928, an investigation into farm costs was conducted for the first time under P.C. Patil (Poona). In this, opportunity costs were calculated and allocated in consonance with the accounting principles avoiding arbitray assumptions. These studies were made to evolve farm cost accounting and income measures suitable for farming in India.

In the Punjab, the Board of Economic Enquiry have been conducting systematic surveys since 1923-24 collecting data for specific purposes. Other surveys which throw some light on farm management have been carried out by Gokhale Institute of Politics and Economic, Poona and the Vishwabharti University, Santiniketan as well as by individual research workers.

More recent and significant from the Methodological point of view is the pilot enquiry into the cost of production of cotton, jowar and groundnut in Akola district in M.P. carried out jointly by the Indian Council of Agricultural Research, Indian Central Cotton Committee and the Indian Central Oilseeds committee in 1953-54 for providing data for technical guidance in the efficient planning of large scale sample surveys (Panse, 1954). From this pilot enquiry, it was concluded that operational holding as against the field should be preferred in future surveys.

1.3 Importance and Consequences of Farm Management Studies

From the foregoing account, it may be seen that the principal objective of most of the investigations has been to calculate the cost of production of crops per unit of area as well as per unit of output and to work out certain income measures. It is only since 1945 that emphasis has been laid more and more on the cost analysis of farms and holdings. The need for maximisation of agricultural output alongwith the economic development of the country has been realised by the Planning Authority and farm management studies will go a long way in helping find a solution to this problem. Moreover, the economic development of the country as envisaged by the First Five Year Plan laid considerable stress on the importance of household economy. Further to raise the standard of living of the growers, it is necessary to take into consideration the existing practices and techniques as are followed and the economic structure of farming as has developed through the past centuries This requires reliable farm management data based on extensive and intensive studies. An analysis of such a data will disclose the existing handicaps and limiting factors which have stood in the way of exploitation of available resources to their full potential. It will thus provide a basis for suggesting measures for rationalising the existing practices to attain maximum efficiency in agricultural production. The farm management data will also help in formulation of sound agricultural policies about land management and land reforms including ceiling on agricultural holdings, if they are broad-based to cover the diverse agricultural practices prevailing in the different regions of the country with their different soil-climate-crop complexes. Besides, the data on costs of production of crops as furnished by these studies will prove helpful in connection with the fixation of prices of agricultural commodities. During and after the war when prices of most of the commodities were controlled, a considerable difficulty was felt in arriving at a fair and equitable price for agricultural producers due to absence of any accurate data regarding cost of the growers producing the particular crop. Even when the prices of most of the major agricultural commodities show a downward trend and a need is felt to stabalise them, the same handicaps stands in the way. The study of cost of production data will enable the cultivator to compare the in put-output relationships of each one of his crops and thus help him in selecting the proper factor-combination. The grower is also able to compare his farm activities with those of his neighbours as also with the overall average as provided by the large scale studies. The cultivator is thus in a position to decide what changes he should bring about to realise better profits.

1.4 Studies in Economics of Farm Management

In view of the importance of reliable farm management data based on extensive studies for the re-organisation and development of agriculture, the Planning Commission, Government of India, and the Directorate of Economics and Statistics in the Ministry of Food and Agriculture sponsored the first series of 'Studies in Economics of Farm Management' in 1954-55 to be undertaken in six typical regions on modern scientific lines. This enquiry was spread over three crop years.

The primary objective of this enquiry was to study the relative merits of the Cost Accounting and the Survey methods of investigations on the one hand and on the other to obtain the experimental data necessary for evolving a scientific methodology in aspects like nature of sampling and nature of analysis to be adopted in future studies. The other objective was to provide information on the input-output relationships in the farms which would be useful for the studies

studies of the relative efficiency of various factor-combinations in the different regions under study. The scheme, in brief, is to co-llect and study data from selected villages; the selection being planned in such a way as to give a broadly representative picture of the production economies of the cultivators in the regions selected.

The scheme has been planned to obtain representative and reliable data from six typical regions in Bombay, Madhya Pradesh, Madras, Punjab, U.P. and West Bengal states. Two contiguous districts from each of these six states have been selected for study in such a way that they represent the most important typical soil-crop-complexes in the state and these regions, taken together, represent all the major cropping patterns in the country. The design of the scheme is the multistage stratified random method of sampling with village as the primary unit and the holding as the ultimate unit. It is to be noted here that holdings are operational holdings and not ownership holdings, meaning thereby the area cultivated by a single operator irrespective of whether it is owned by him or taken on lease. Each district was divided into two homogeneous zones or strata and from each zone villages were selected with probability proportional to the cultivating population of the village. Both the methods of investigation viz. Cost Accounting and Survey were employed in all the states except Madhya Pradesh where only the Cost Accounting method was adopted. In each zone, the sample consists of five villages for each method of which two villages are common to both. Thus, in effect, eight villages have been selected for investigation from each zone. There are, therefore, 32 villages selected from each state. The next stage of sampling is that of selection of holdings within each selected village. Under the scheme, 10 holdings have been selected for the Cost Accounting method and 20 for the Survey method. For the purpose of selection of holdings, all the holdings in a village have been ordered according to the size of holding in the year previous to the enquiry. The total number of holdings are then divided into five groups each group having equal number of holdings. Two holdings have been selected at random from each group for the Cost Accounting method and four for the Survey. In common villages, precautions were taken to exclude those holdings which have once been selected for the Cost Accounting method from being included again under the Survey method. The total sample thus consists of 600 holdings in each state, 200 in the Cost Accounting sample and 400 in the Survey.

The selection of villages was done by the Directorate of Economics and Statistics in the Ministry of Food and Agriculture and that of holdings by the respective officers-in-charge of the scheme in each region.

In U.P., however, as the population figures for the villages were not available, the selection of villages was made with probability proportional to the total area of the village.

1.5 <u>A Description of the Two Methods of Collecting the Data</u>

As mentioned earlier, for investigation into the farm management research, two principal methods of collecting data viz.

(i) Cost Accounting

and (ii) Survey

are generally recognised as the most suitable ones. The former is characteristic of an intensive study of farm management practices and involves the maintainance of daily records of all operations and business transactions conducted on the farm in details. The system affords highly reliable data from the research point of view if the records are kept properly and accurately. The Survey method which is also called the extensive method, does not demand such minute details. It involves a number of visits over a larger area and recording of information as reported by the farmers. The chief drawbacks in the

Cost Accounting method are that it is time-consuming and far more costly than the other method. The Survey method on the other hand is comparatively cheap and yields quick results but its drawbacks are that the estimates are likely to be less reliable and it may not provide all the detailed data required for studying certain types of farm management problems. Thus both the methods have their own advantages and limitations. To get an idea of the difference of cost in the two methods, the expenses on the field-staff in the year 1954-55 for the Cost Accounting and Survey methods were of the order of Rs.37900 and Rs. 11000 respectively; the difference being Rs. 26000. In the present enquiry both the methods have been employed to guage the reliability and efficiency of Survey method by a comparis-on of the data collected by this method with that collected by the Cost Accounting method from the same population.

The present study based on the data collected under the scheme 'Studies in Economics of Farm Management' concerns:

(i) The input-output relationships, with a view to determine the function which could efficiently express this relationship and could be used as a prediction function,

(ii) The comparison of the two methods of collecting the primary data viz. the Cost Accounting and the Survey with a view to guage the suitability of the latter,

(iii) Determination of the size of the sample viz. the number of villages and holdings to be sampled in such investigations for given precision,

(iv) The study of the efficiency of stratification within the villages as adopted in the scheme.

2 REVIEW OF LITERATURE

2.1 A great amount of research work has been done in other countries, the maximum contribution being from the U.S.A. A good many books and papers in bulletins and journal have been written on different aspects and problems of farm management. We are here chiefly concerned with the analysis of farm management data. In what follows is given a review of the available literature dealing with this.

2.2 Defining farm management, Heady (1952) writes:

The goals of agricultural production economics are two fold:

(i) to provide guidance to individual farmers in using their resources most efficiently, and

(ii) to facilitate the most efficient use of resources from the standpoint of consumer's economy.

Now economics deals with the choice between the alternatives. Problems of choice arise only when resources are limited and alternative uses can be made of them. The field of economics deals with means and ends. Ends or objectives may deal with profits, consumer's satisfaction or physical production during wartime. Means are concerned with the physical resources, funds, organisation etc. which can be used to attain the various possible objectives. As a science of choice between alternatives, economics is also concerned with the choice of goods and services which define the conditions of maximum utility and satisfaction of consumers i.e. it is concerned with the conditions which are necessary if a given amount of product or profit is to be produced with a minimum of costs or resources.

There are two major problems in economics:

(i) One problem concerns the organisation of production. The problem of choice in production is termed resource allocation, i.e. how resources should be allocated between different products or production methods at a certain point of time or in different periods of time.

(ii) Second overall problem is that of organisation of consumption which may be termed as income allocation viz. the allocation of income or product between different alternatives at a certain point of time or over a period of time.

The laws of production explain the conditions under which quantities can be maximised (profit, physical output, national income etc.) or minimised (cost, physical input etc.).

Thus the objectives of the agricultural production economist are:

(i) to determine and outline the conditions which give the optimum use of capital, land, labour and management in the production of crops and livestock,

(ii) to determine the extent to which the existing use of resources deviates from the optimum,

(iii) to analyse the factors which condition production patterns and resource use and

(iv) to explain means and methods in getting from the existing to optimum use of resources.

The phase of agricultural production economics dealing with intra-farm allocation or use of resources has been termed 'farm management'.

Describing the laws of return and nature of input-output relationship, Heady writes:

The relationship between input of a single variable factor (with the quantities of other resources held constant) and output of a single product is one of the three general forms. These are

(1) constant productivity,

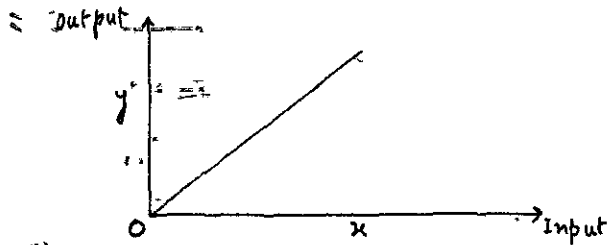
(2) decreasing productivity and

(3) increasing productivity.

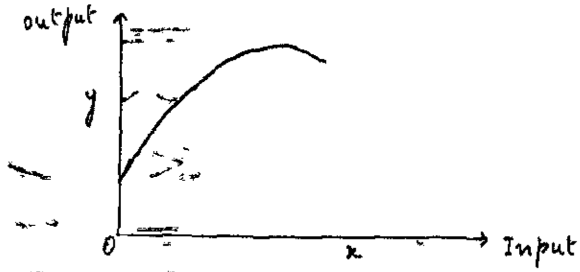
Constant or increasing productivity is seldom found when only one factor is varied and all others kept constant.

(1) Constant returns: This holds true if all units of the variable input factor which are applied to the fixed factor result in equal

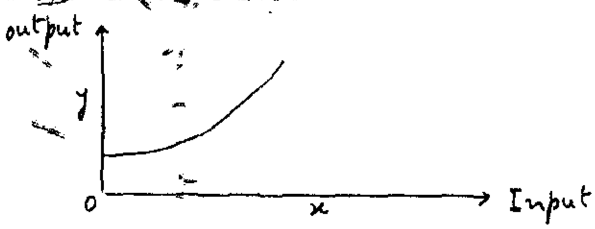
additions to the total output of the product i.e. the input-output relationship is linear.



(2) Diminishing returns: This holds when each additional unit of input adds less to the total output than the previous unit. The relationship is not linear as the total product is increasing at a decreasing rate.



(3) Increasing return: This holds when each successive unit of input adds more to the total product than the previous unit i.e. the total product increases at an increasing rate.



Production functions are possible which combine any two of the properties outlined above e.g. increasing-decreasing functions illustrated by

$y = a + bx + cx^2 + dx^3$

where a, b, c and d are positive constants.

To explain the above principles mathematically, we need define the terms average and marginal products or costs. Average product or cost is the ratio of total product or cost divided by the total variable input used.

i.e. ~~marginal~~ Average product or cost = y/x

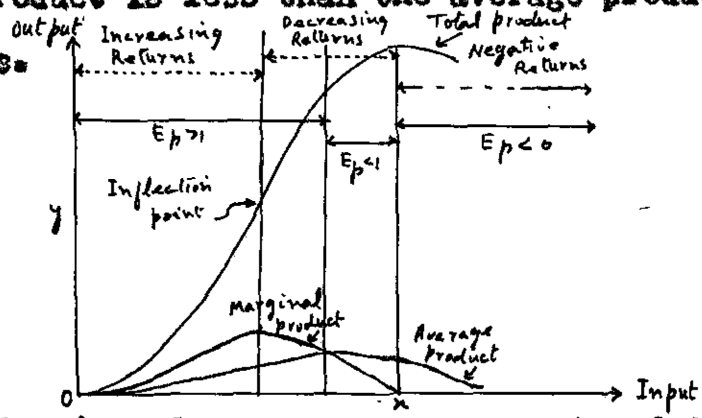
Marginal product or cost is the ratio of the increase (say Δy) in the total output or cost corresponding to an increase (say Δx) in the variable factor input.

i.e. marginal product or cost = $\Delta y / \Delta x$

Thus, as long as $\Delta y/\Delta x$ is increasing, the total product must increase at an increasing rate, with $\Delta y/\Delta x$ decreasing but positive, the total product must increase at a decreasing rate and at $\Delta y/\Delta x$ constant, total product must increase at a constant rate.

We are now in a position to describe the widely accepted principl known as the law of diminishing returns:

If the quantity of one productive service is increased by equal increments with the quantities of other resource services held constant, the increments to the total product may increase at first, but will decrease at a certain point. The point referred to is the maximum point on the marginal product curve which is also the point of inflection on the total product curve. Similarly as long as the marginal product is greater than the average product, the average product increases; if the marginal product is less than the average product, the average product decreases.



2.3 Bradford and Johnson (1953) while defining management and its functions, write :

Management is an intangible part of production which develops within the lives of men. Management functions when a farmer is

(i) observing and conceiving ideas,

(ii) analysing with further observations,

(iii) making decision on the basis of analysis,

(iv) taking actions and

(v) accepting responsibilities.

Managerial ability is partly inherited and partly acquired but efficient management can be and must be acquired for efficient running of farm. On the question as to how we measure managerial capacity, a number of tests have been devised by industrial psychologists.

Defining farm business and allied enterprises, the authors write:

The physical parts of farm business are

(1) Land,

(2) Improvement including such items as buildings, fences, and water systems,

(3) Labourers,

(4) Machines and other types of equipment,

(5) Feeds, fertilisers and seed etc.,

(6) Usually productive livestock,

(7) Money for operating and

(8) The farm family with its wants, preferences, and desires.

The other aspects or parts of farm businesss are

(a) Crops produced,

(b) Receipts which may be defined to include

(i) the cash value of items sold,

(ii) the value of farm products produced by farm business and used by the farm families,

(iii) the value to the operator of the use of the house, personal consumption of a part of the farm real estate investments.

The last aspect concerns the expenses. These are of four types:

The first type involves items which are used within the year e.g. costs of feed, fertilisers purchased, labour, fuel, repairs, taxes, insurance etc.

The second type involves depreciation of properties i.e. repairs and maintainance of buildings and machines not effected by cash expenses.

The third type involves decreases in inventories of operating capital, feed and supplies other than feed. These may be called other inventory decreases.

The fourth type involves use of labour services performed by family members who are not paid for in product or in cash.

Thus we have four kinds of expenses:

(i) Cash items i.e. the current expenses,

(ii) Depreciation of physical properties,

(iii) Other inventory decreases, and

(iv) Money value of family labour.

2.4 Redman (1954) has discussed the problems and possible solutions in determining input-output relationships in agricultural enterprises. He writes:

The importance of input-output relationships cannot be denied by any economist. In establishing this, some difficulties have to be faced viz.

(i) those involving human aspects,

(ii) those arising out of the nature of existing data,

(iii) those encountered in securing additional data, and

(iv) those encountered in using the existing data and additional data.

(i) Large errors in estimating input-output coefficients may result from personal bias of the research worker in selecting and using information. Personal bias may also result in the method used. Similarly fear of offending established dogmas and respected sentiments is a strong hindrance to research in agricultural economics. In early agricultural research, it often was expedient to use an intutive idea or to make simplifying assumptions. But when such assumptions as linear relationship in livestock feeding etc. exist for long period, it becomes a problem to getting research workers challenge the accepted ones, design new experiments

experiments to obtain new evidence and incorporate the findings into a logical framework to which future work can be added.

(ii) The economist is interested in such problems as which combination of inputs is achievable and which is not; how much product can be obtained from a given combination of factors or how much of one input would be required to produce a specific amount of product, given the combination of other inputs, so that he can determine the most efficient combination or choice. Many problems are involved in using the current data for input-output work e.g. many biological experiments consider no alternatives from which a farmer can choose because they have been designed to compare a single point on a given production with a single point on another function e.g. water vs nitrogen, boron vs potash etc. These experiments are of value for certain purpose but they are distinctly limited for use in establishing input-output relationship. Problems result also when response data to variable inputs are obtained by assuming the level of fixed inputs constant when actually they are not. The resources obtained from an input depends in a large part on the level at which other inputs are fixed as well as the number of units of the variable input already used e.g. in supplemental irrigation, the response to nitrogen as compared to no irrigation depends upon moisture level already in the soil. Similarly the soil fertility can be considered.

(iii) Thus much of the physical data available from past experimentation are not suited for determining adequate or logical input-output relationships and when used, extreme care must be exercised. Thus arises the need for additional data which should be collected after designing according to requirements and conducting future experiments properly. The sample survey as a means of securing the data has been used from the very beginning. It is essential in using certain methods to develop a model or hypothesis of presumed relationship to be examined and avoid gathering the data on mixed

production functions. Similarly for the method of Cost Accounting.
(iv) However good raw data one has got unless utilised accurately and fully, much value will be lost. The synthetic method of handling the data lends itself to the construction of cost functions which, to be meaningful and useful, must be identified with a specific production process, planning span and level of prices. The surveys and techniques should also be improved in the light of the improvements made in the recent years.

2.5 Zvi Griliches has discussed the specification bias in estimates of production functions. He writes:

It is common in empirical work to compromise and use second best methods or variables. Due to limitations, we exclude variables, accept approximations and commit various other sins of omission and commission called the 'specification errors'. Knowing their effects upon our results, we may be better able to interpret them. In what follows a method for ascertaining the consequences of some of those compromises will be described.

In the Cobb-Douglas production function

$$y = a_1 \log x_1 + a_2 \log x_2 + \ldots\ldots + a_k \log x_k ,$$

a_i is the elasticity of production of the i-th variable. Let e

$$e = \sum_{i=1}^{k} a_i$$

Then e is called the elasticity of production (of the whole function) or the function coefficient. It will be shown that we underestimate returns to scale if we exclude an input that varies more than proportionally with the included inputs and vice versa.

<u>The Omission of Relevant Variable</u>

If we omit a relevant variable which is uncorrelated with the other independent variables in our analysis, the omission will not bias the estimates of the parameters of the included variables. But even in this case we shall have a biased estimate of e.

Under the usual assumption about the coefficients of production functions a_i (a_i 0), our estimate of e will be biased downwards and the bias will be equal to minus the coefficient of the excluded variable. For let x_k be the excluded variable; if the b's are the estimated coefficients and if $\hat{e}$ is the estimate of e, then

$$\hat{e} = b_i \quad \text{and} \quad \text{bias} = E(\hat{e} - e)$$
$$= E\left[\sum_{i=1}^{k-1} b_i - \sum_{i=1}^{k} a_i\right]$$
$$= \sum_{i=1}^{k-1} a_i - \sum_{i=1}^{k} a_i$$
$$= -a_k .$$

However the assumption of no correlation between the excluded variable and all the included variables is hardly tenable for factors of production . A more plausible assumption is that there is a positive correlation between the excluded variable and some of the included variables. This will bias the estimates of at least one of the coefficients upwards but it is not clear if it will be sufficient to compensate for the downward bias in the estimate of e due to the omission of the coefficient of the excluded variable. The expectation of the estimates of the included coefficients is given by

$$E(b_i) = a_i + p_{ik}a_k$$

where in the case of Cobb-Douglas function, the p's come from the following regression:

$$\log x_k = p_{1k}\log x_1 + \ldots\ldots\ldots + p_{k-1}\log x_{k-1} + V$$

i.e. the p's are the coefficients in the regression of the logarithm of the excluded variable on the logarithm of all the included variables. The sign of p_{ik} will determine if a particular b_i will underestimate or overestimate the corresponding true parameter a_i (assuming $a_i > 0$). The sign of p_{ik} will depend on the actual distribution of all the x's in a particular sample. Although a negative association of x_k with some of the x's is possible, for factors of production a positive association is more plausible. Hence if we exclude some factor of production from our regression, we shall on the average overestimate at least some of the coefficients of the included variables. However we are interested in what happens to our estimate of returns to scale

$$\hat{e} = \sum_{i=1}^{k-1} b_i ,$$

the sum of the estimated coefficients, as compared with the true returns to scale

$$e = \sum_{i=1}^{k} a_i ,$$

the sum of the coefficients in the true production. Now the bias in our estimate of returns to scale will be equal to

$$\begin{aligned} E(\hat{e} - e) &= E\left[\sum_{i=1}^{k-1} b_i - \sum_{i=1}^{k} a_i\right] \\ &= \sum_{i=1}^{k-1}(a_i + p_{ik}a_k) - \sum_{i=1}^{k} a_i \\ &= \sum_{i=1}^{k-1} a_i + a_k\sum_{i=1}^{k-1} p_{ik} - \sum_{i=1}^{k} a_i \\ &= a_k(\sum p_{ik} - 1) \end{aligned}$$

That is, whether we underestimate or overestimate returns to scale will depend on the sum of the coefficients in the 'auxiliary' regression. We shall underestimate e if this sum is less than unity, there will be no bias if it equals unity and we shall overestimate e if this sum is larger than unity.

i.e. if $\sum p_{ik} \lesseqgtr 1$, then bias $\lesseqgtr 0$.

A very simple interpretation of this sum of coefficients is possible in the 'auxiliary' regression. These are coefficients in the following Cobb-Douglas function:

$$x_k = V x_1^{p_{1k}} x_2^{p_{2k}} \cdots x_{k-1}^{p_{k-1,k}} .$$

The sum of p's is equivalent to the elasticity of production in this particular 'auxiliary' regression. The bias in our estimate of e will depend on how x_k changes in the sample when all the other x's are varied to scale. We shall underestimate e if proportional changes in the included variables are associated with less than proportinal changes in the excluded variable in our sample. We shall overestimate e if the excluded variable varies more than proportioanally with the included variables. There will be no bias if the excluded variable varies on the average in the same proportion with proportinal variation in all the included variables. Similarly if we leave out m relevant variables while including l, the expectation of our estimated coefficients will be given by

$$E(b_i) = a_i + \sum_{k=l+1}^{l+m} p_{ik} a_k$$

The bias in our estimate of e is given by

$$E(\hat{e} - e) = \sum_{k=l+1}^{l+m} a_k \left(\sum_{i=1}^{l} p_{ik} - 1\right).$$

The sign of the bias in our estimate of e will now depend not only on whether the excluded variables change more or less proportionally with the included variables but also on the relative magnitudes of the coefficients attached to the variables that change less proportionally with the included variables.

3 MATERIAL FOR STUDY

3.1 The data actually used related to the 'Studies in Economics of Farm Management' and were obtained from the Directorate of Economics and Statistics, Ministry of Food and Agriculture. These data belong to the districts of Meerut and Muzaffarnagar in U.P. and refer to the first year of the enquiry viz. 1954-55.

Other data used for the purpose of fitting a comparable input-output relationship was obtained from the Indian Central Sugarcane Committee for Sugarcane crop only and refer to year 1955-56.

3.2 These data have been analysed, not primarily to comment on the agro-economic results of the investigations, But as illustrative material for studying some aspects of the statistical methodology involved in these investigations. The particular data were taken up for analysis as being most readily available in sufficient details.

The available data contained information in four crops viz. Wheat, Sugarcane (planted), Sugarcane (ratoon) and Gram. Out of these, three crops Wheat, Sugarcane (planted) and Gram were chosen for analysis and other studies.

4. Input-Output Relationships

4.1 This section concerns the fitting of appropriate production furnction to the data. Various production functions have been tried and fitted to the data on different items. The functions which are usually fitted on the farm-management data, are the following :

(i) Spillman's production function.

This function is of the type

$$y = m - ar^{x}$$

where y is the output,

x is the variable resource input,

m is the maximum attainable output,

a is the maximum addition due to the variable input,

and r is the ratio of the increment in the output for a small increment in the variable input to the succeeding increment in the output for the same increment in the output.

The elasticity of production, E_p, is defined as the ratio of the percentage increase in the output to the percentage increase in the input. It varies in the above function, being given by

$$E_p = \frac{\Delta y}{y} / \frac{\Delta x}{x}$$

In the case where the law of diminishing returns holds and $r < 1$ with the increase in x, y increases at a decreasing rate, so that when x is very large, y approaches its maximum value m.

In the above function y never decreases so that we do not have negative returns for addition in variable input (Heady, 1952). Thus this function may not be appropriate to the data or to that range of observations in which negative returns may be occuring.

Note: x and y will denote input and output respectively throughout this section.

(ii) Cobb-Douglas production function.

This is the most commonly used production function and is of the form

$$y = ax^{b}$$

where a is the output for unit input

and b is the constant elasticity of production

In the case of more than one (say n) variable input, the function may be written as

$$y = a \prod_{i=1}^{n} x_i^{b_i}$$

where b_i is the constant elasticity of production of the i-th variable input

and a is the output when one unit of each of the variable input is used.

We have constant returns if the elasticity of production or the sum of the elasticities in the case of more than one variable input is unity. A 1% increase in the variable input (or inputs) gives rise to a 1% increase in the output. When $E_p > 1$, we have increasing returns for each additional unit of input (or inputs) and when $E_p < 1$, we have diminishing returns for each successive unit of input (or inputs). The same holds true for each individual input in case we use more than one variable input. (Heady, 1952; Bradford and Johnson, 1953).

The same drawback as that encountered in the Spillman's production function is present in the Cobb-Douglas production function viz. it does not allow a negative marginal product(under the assumption that the elasticity of production is > 0). Another limitation of the Cobb-Douglas production function is that it does not place any upper limit on the total output in the manner of the Spillman's production function. Thus it should not be used beyond the range of observations included in the original data.

(iii) In view of the inadmissibility of negative marginal returns, both the functions described in the previous paragraphs are less acceptable than the simple polynomial of the form :

$$y = a + bx + cx^2$$

which is the common quadratic function, a, b and c being constants. Since this function allows negative marginal product, it would be useful to try it on data in which the initial additions in the input give increasing marginal returns but further additions give constant and later negative returns.

(iv) Another form of production function, known as the quadratic squareroot function, is of the form :

$$y = a + b\sqrt{x} + cx$$

where a, b and c are constants.

This function also admits negative marginal returns and behaves similar to the quadratic type. The only difference between this function and the quadratic type is that the variations may be considerably reduced in the former.

Of all the four functions so far described there is one advantage in the Cobb-Douglas production function viz. that we may analyse the data for upto 5 or 6 variable inputs with reasonable accuracy and still avoid heavy and complicated computations as it becomes linear on the logarithmic scale. We cannot consider more than three variable inputs in the quadratic or the quadratic square-root type and in the case of Spillman's production function it would be difficult to go in for more than one variable input.

(v) The fifth and the easiest of all to fit is the well known linear function viz.

$$y = a + bx$$

but it does not reflect any increasing or decreasing returns but just gives constant marginal returns. It will not show any change in the trend of output.

4.2 Choice of Production Function

For the analysis of the data we have to choose from among the various production functions outlined above. Now the most common and widely used production function for such purposes is that of the Cobb-Douglas type. Also by drawing scatter diagrams between the output / acre and the various items of input / acre (enumerated later) on the original scale as well as on the logarithmic scale and comparing the corresponding ones, it was found that the variability reduced to some extent in the latter case. In some cases quadratic functions were also fitted. To recall the chief properties of the Cobb-Douglas production function, we have increasing marginal returns to all the inputs used jointly if the sum of the elasticities of production is greater than unity, decreasing marginal returns if this sum is less than unity and constant marginal returns if it is equal to unity. Further that this function does not admit of negative marginal returns (under the assumption that each of the elasticities of production is greater than zero) and that it does not put any upper limit to the total output. Thus we should not use this function beyond the range of observations included in the original data.

Scatter diagrams on the original scale and on the logarithmic scale were drawn for the following cases:

Wheat (irrigated): Total Output/acre vs Human Labour/acre, Bullock Labour/acre, Seed/acre, Irrigation charges/acre, Fertilisers and Manures/acre and Total Input/acre.

Sugarcane (planted, irrigated): Total Output/acre vs Human Labour/acre and Bullock Labour/acre

Wheat (unirrigated): Total Output/acre, vs. Human Labour/acre Bullock Labour/acre, Fertilizers and Manures/acre

Gram (unirrigated): Total Output/acre vs Human Labour/acre, Bullock Labour/acre, Seed/acre and Land Revenue/acre.

There was no irrigated Gram crop.

All the above observations were taken in terms of their money value (i.e. in rupees).

4.3 The method of fitting the production function was that of the least squares method i.e. error s.s. was minimised. The Cobb-Douglas production function reduces to the ordinary linear function on the logarithmic scale. Thus we have

$$\log y = a + \sum_{i=1}^{n} b_i \log x_i$$

if we have n variable inputs.

Denoting the variables by capital letters on the logarithmic scale, we have

$$Y = a + \sum_{i=1}^{n} b_i X_i$$

The error s.s. $\sum\left[Y - \left\{a + \sum_{i=1}^{n} b_i X_i\right\}\right]^2$ is minimised.

Thus differentiating, we obtain the normal equations of the type $\sum X_j Y = a \sum X_j + \sum_{i=1}^{n} b_i \sum X_i X_j$ $(j = 1, 2, \ldots\ldots, n)$

Solving those n normal equations we get the values of b's.

Taking X's to be normally distributed (the number of observations in each of the X's being of the order of 150) we test the partial regression coefficients b's.

4.4 The variables in the production functions fitted to the various crops are as follows :-

y stands for	Output/acre	(money value in rupees).
x_h " "	Human Labour/acre	(-do-)
x_b " "	Bullock Labour/acre	(-do-)
x_s " "	Seed/acre	(-do-)
x_i " "	Irrigation charges/acre	(-do-)
x_f " "	Fertilisers and Manures/acre	(-do-)
x_l " "	Land Revenue/acre	(-do-)

Four or more of the above variable inputs have been employed in the various production functions that follow.

(i) Wheat (irrigated).

The function fitted is

4.4.1 $$y = 1.8559\, x_h^{0.0559}\, x_b^{0.0409}\, x_s^{0.2139}\, x_i^{0.0627}$$

The coefficient of x_s viz. Seed/acre was alone found to be significant at 5% level. The coefficient of determination, R^2, is given by

$R^2 = 0.0632.$

Therefore the multiple correlation coefficient, R, is given by

$R = 0.2521$ Not significant.

The sum of the elasticities of production is given by

$E_w = 0.3734$

As the individual elasticities as also the overall elasticity of production are each less than unity, the function indicates decreasing marginal returns to all the inputs used individually as well as jointly.

Other functions fitted to the individual variable inputs in the case of Wheat (irrigated) crop are

(i) Linear functions,

(ii) Quadratic square root functions.

(a) For Human Labour/acre (money value) we have

4.4.1 (i) $y = 196.67 + 0.5\, x_h$

The regression coefficient is not significant.

4.4.1 (ii) $y = 196.84 - 0.0503\sqrt{x_h} + 0.5036\, x_h$

Both the regression coefficients are not significant.

(b) For Bullock Labour/acre (money value), we have

4.4.1(iii) $y = 206.92 + 0.1007\, x_b$

The regression coefficient is not significant.

4.4.1 (iv) $y = 206.00 - 0.918\sqrt{x_b} + 0.0972\, x_b$

Both the regression coefficients are not significant.

(c) For Seed/acre (money value), we have

4.4.1 (v) $y = 176.50 + 3.5162\, x_s$

The regression coefficient is significant at 5% level.

4.4.1 (vi) $y = 176.00 + 0.040\sqrt{x_s} + 3.5157\, x_s$

Only the partial regression coefficient of x_s is significant at 5% level.

Thus, of all the above three variable inputs, only the Seed factor contributes significantly to the output, a result in conformity with that derived from the previous Cobb-Douglas production function.

(ii) <u>Sugarcane (planted).</u>

The function fitted is

4.4.2 $y = 2.2121\, x_h^{-0.0278}\, x_b^{0.0239}\, x_s^{-0.0369}\, x_i^{0.2094}\, x_f^{0.1525}$

The elasticities of production due to the variable inputs of Human Labour/acre and Seed/acre are negative whereas those of the other three are positive. None of these is significant. The coefficient of determination is given by

$$R^2 = 0.0319$$

and therefore the coefficient of multiple correlation is given by

$$R = 0.1786 \quad \text{Not significant}$$

The sum of the elasticities of production is given by

E_s = 0.3218

As the individual elasticities as also the overall elasticity of production are each less than unity, the function indicates decreasing marginal returns to all the inputs used individually as well as jointly.

(iii) Gram

The function fitted is

4.4.3 $$y = 1.016\ x_h^{0.165}\ x_b^{0.210}\ x_s^{0.407}\ x_1^{0.039}$$

The first three coefficients are significant at 5% level.

The coefficient of determination is given by

R^2 = 0.128

and the coefficient of multiple correlation is given by

R = 0.357 Significant at 1% level.

The sum of the elasticities of production is given by

E_s = 0.821

As the individual elasticities as also the overall elasticity of production are each less than unity, the function indicates decreasing marginal returns to all the inputs used individually as well as jointly.

4.5 From the foregoing function we see that only a few factors contribute significantly to the production of the three major crops. The coefficient of multiple correlation in two of the three crops are not significant (that of the third crop although significant accounts for only 13% of the variation), showing that the production functions for these crops are of little practical value in explaining the variation in output or predicting it. The reason for this may be that

(i) these being the data collected in the first year of the scheme, sufficient number of adequately trained personnel was not available and therefore the data may not be reliable i.e. the quality of the data is poor, or

(ii) there may really exist no appropriate functional relationship between the various input and output values.

Before taking any of the above two reasons or explanations as basis of our conclusions, there is a need to augment our information with similar study of more data collected from various sources.

4.6 Data on Sugarcane (planted as well as ratoon) was also collected by the Indian Central Sugarcane Committee from U.P. The data belonging to the same region, as in the Farm Management Enquiry were obtained from the above mentioned source. These data were also analysed so as to obtain additional information on the Sugarcane crop.

The function fitted is

4.6.1 $$y = 0.8709\ x_h^{0.3166} x_b^{-0.0788} x_s^{0.6233} x_f^{0.553} x_i^{0.0216}$$

(Here y stands for Total Output/acre in maunds)

The coefficient of x_h and x_s viz. Human Labour/acre and Seed/acre are significant whereas that of x_f viz. Fertilisers and Manures/acre is very near the significance level.

The coefficient of determination is given by

$$R^2 = 0.2869$$

and, therefore, the coefficient of multiple correlation is given by

$$R = 0.5356$$ Significant at 1% level.

The sum of the elasticities of production is given by

$$E_s' = 0.934$$

As the individual elasticities as also the overall elasticity of production are each less than unity, the function indicates decreasing marginal returns to all the inputs used individually as well as jointly.

Observing that the two inputs viz. Human Labour/acre and Seed/acre contribute significantly to the production, we fit a quadratic function for the above two inputs.

The function thus fitted is

4.6.2 $$y = 0.5803 + 1.9843\ x_h + 3.5570\ x_s - 0.0499\ x_h^2 - 0.1792\ x_s^2 + 0.1542\ x_h x_s$$

None of the partial regression coefficients is significant.

The coefficient of determination is given by

$R^2 = 0.3514$

and, therefore, the coefficient of multiple correlation is given by

$R = 0.593$ Significant at 1% level.

As the coefficient due to the variable input x_f viz. Fertilisers and Manures/acre was nearly significant, we add the same to the above quadratic function.

The function then becomes

4.6.3 $$y = 16.2389 + 1.3694\ x_h - 0.4310\ x_s + 1.1550\ x_f + 0.0158\ x^2_h + 0.1608\ x_s^2 + 0.0720 + 0.0720x^2 + 0.0803\ x_h x_s + 0.0788 x_h x_f$$

$x_s\ x_f$ has not been included as a variable as the correlation between x_s and x_f was found to be not significant.

The coefficient of determination is given by

$R^2 = 0.6304$ Significant at 1% level.

4.7 Comparing the two production functions fitted to the data on sugarcane crop obtained from two different sources, we observe that the first one viz. that on the data of 'Studies in Economics of Farm Management' accounts for a small percentage of the variation whereas that on data of Indian Central Sugarcane Committee accounts for a reasonably good percentage of that variation. A similar impression on other crops would be very interesting.

The present comparison indicates that the possibility of the farm management data for the first year not being of the required standard of quality cannot be ignored, since under similar conditions, data collected by experienced and better trained field-staff have led to a distinctly closer relationship between output of sugarcane and various other input factors. Another important conclusion that emerges from the comparison is that the input factors included in the study actually contribute to the production only to a limited extent but that there are other 'extraneous' factors, which cannot be controlled, that are responsible for influencing the production. Consequently even the improved production function leaves a major part of the variation in output unaccounted for and thus not effective for prediction. In any future study of this type, it would be worth while to find out and study the effect of these 'extraneous' factors e.g. climate, soil-fertility etc. Thus if we can divide the land according to the fertility-gradient and then try to fit appropriate production function, we may get one that accounts for most of the variation in production.

....

5. A Comparison Of The Cost Accounting And Survey Methods Of Collecting The Primary Data

5.1 As already stated, the Cost Accounting method is an intensive method in that detailed records of day to day business on the farm are maintained. The data so collected can be very reliable and complete in all respects. The Survey method is an extensive method in that we just rely on the farmer for suppling us the required information periodically. The farmer might give us wrong information partly because he does not remember the expenses incurred by him long back and partly because he might not wish to give the correct information. Another consideration and a very important one is that of the cost involved. Obviously the cost of the former method will be more than that of the latter. Thus a comparison between the two approaches becomes a very important factor. In case we find that the two methods do not differ appreciably in the quality and completeness of information provided, it would be worthwhile to adopt the Survey method in future investigations as we would be saving a handsome amount of money. But in case the two methods differ significantly it would be better to adopt the Cost Accounting method in spite of the heavy cost involved, because if we adopt the Survey method the date so collected will not be reliable and therefore the conclusions and recommendations made on the basis of this data will be of little value. It will, therefore, not be a saving of expenses but a mere waste of money and time.

The difference in the working of the two methods is as follows : one fieldman under the Cost Accounting method collects data from 10 farms in one village only throughout the year, while the investigator under the Survey method collects data from 100 farms in five different villages situated at long distances from one another. The two methods differ particularly in respect of the visits paid to the farmers by the fieldmen and investigators. In the Cost Accounting method, the fieldman meets each farmer at least twice or thrice a week and can increase his frequency whenever necessary and can check up any doubtful data within a very short time. In the Survey method, on the other hand, the farmers of each of the five villages are approached four times a year in rotation. Any data subsequently found to doubtful will have to wait at least three months before they are checked up and corrected. This time lag takes away much of the reliability and value even of the corrected data. Moreover, the data collected within a week at the latest after the operations are completed as is done under the Cost Accounting method, are likely to be more reliable than the information collected under the Survey method long after the completion of operations.

To recall relevent details of collection of data and selection of villages and holdings, in each zone five villages were selected for each method of which two were common. Thus effectively we have eight villages in all from each zone i.e. 32 villages constitute the entire sample. The comparison is made in two parts. Firstly that based on independent villages by comparing the village means with appropriate standard errors. The second is based on common villages by comparing group means. As already stated, the holdings were first ordered and divided into five equal groups. Two holdings for the Cost Accounting method and four for the Survey method were selected from each group. Due to some accidents or non-response from the farmer, observations from certain holdings could not be recorded in some groups thus resulting in disproportionate cell frequencies. In a common village, therefore, the approach of two-way classification with unequal cell frequencies has been followed. (Kendall, 1951; Rao, 1952; Das, 1953).

5.2 We first deal with the type of comparison based on independent villages. Two sets of tables have been given here. In the first set of tables, the weights given to various strata are proportional to the strata sizes. This table will be valid for use only in those regions where the strata are defined in a way similar to the present ones. But to get an idea of the results irrespective of the effects of variation in strata sizes, equal weights for the various strata have been used in the second set of tables. However a look at the two sets of tables reveals little difference in the results and therefore either of the two may be selected for further use.

As is expected, the results with equal weightage have generally slightly lower sampling errors. If we consider the present investigators as a random sample of investigators, the second set of tables is more appropriate whereas if we wish to take into account the variations in the cultivating population in the various strata, the first set of tables is more appropriate in so far as the results are influenced by the biased information given by the farmers in spite of the best efforts of the investigators. The second type of comparison is the one based on common villages. The approach of two-way classification with unequal cell frequencies, followed in a common village is as follows :-

Assuming no interaction between the methods and groups the sum of squares (s.s.) for the methods (adjusted for the disproportionateness in cell frequencies) and the s.s. for the groups are calculated. Interaction s.s. is got by subtraction. The interaction means square (m.s.) is then tested against the error m.s. If it is not significant, it is pooled with the error m.s. to get a better estimate of the error m.s. The adjusted m.s. for the methods is then tested against the pooled error m.s. If on the other hand, the interaction m.s. is significant, the s.s. for the methods is recalculated (now on the basis of assuming the interaction to be present) and tested against error m.s. As in the type of comparison between the two methods in the case of independent villages, t-test has been employed, the above method will also be reduced to t-test form remembering that $F_{(1,n)} = t^2_{(n)}$. This procedure will be amply clear when theoretical details have been given illustrated with a few examples dealt with in details.

(Kindly read page 42 before page 41)

5.8 We now come to the other type of comparison viz. that based on common villages. In a common village, the holdings selected for the Cost Accounting method were different from those selected for the Survey method. As we have five groups from which holdings were selected independently, the variation in the holdings may be due to (i) methods or (ii) groups. Another source of variation may be that the effect of methods differs from group to group or vice versa i.e. there may be interaction between the methods and the groups. So we have three sources of variation viz. that due to (i) methods, (ii) groups and (iii) interaction between the first two; so that if we have q groups, the analysis of variance table for a village will be as follows :-

Source of Variation	Degrees of Freedom
Methods	1
Groups	q-1
Interaction	q-1
Error	n-2q
Total	n-1

where n denotes the total number of observations in all the cells.

As the cell frequencies are different, we adjust the s.s. due to methods for the disproportionateness in cell frequencies and on the hypothesis that the methods do not differ, we test the methods m.s. (adj.) against the appropriate error m.s. The theoretical details are as follows :

Taking first the two-way classification with unequal cell-frequencies (Das, 1953; Kendall, 1951), let the two factors be denoted by A and B with levels, a_1, a_2, ..., a_p and b_1, b_2, ..., b_q.

A comparison of the Cost Accounting and Survey methods of collecting data based on independent villages.

Let S denote the group mean of the observations under Survey method
and C " " " " " " " Cost Accounting method.

Table 5.1

	Proportional Weights				Equal Weights			
Wheat:								
Factor	(S-C)	S.E.(S-C)	d.f.	t	(S-C)	S.E.(S-C)	d.f.	t
H.L.	2.509	3.4464	8	N.S.	1.877	3.8708	8	N.S.
B.L.	70.768	13.1976	8	**	73.658	14.2441	8	**
Seed	0.680	0.7618	8	N.S.	0.452	0.7245	8	N.S.
T.Input	77.429	16.3098	8	**	79.799	18.3306	8	**
T.Output	14.456	16.8365	8	N.S.	14.310	15.8328	8	N.S.
Sugarcane (planted):								
H.L.	0.456	11.1292	8	N.S.	4.179	10.8706	8	N.S.
B.L.	48.924	16.2876	8	**	55.429	14.8000	8	**
Seed	0.402	5.6991	8	N.S.	2.400	5.0275	8	N.S.
F.&M.	20.661	7.0352	8	**	16.419	6.0880	8	**
T.Input	26.529	25.6307	8	N.S.	41.747	23.0187	8	N.S.
T.Output	30.440	62.6895	8	N.S.	31.937	54.8575	8	N.S.
Gram:								
H.L.	3.103	1.8690	8	N.S.	3.037	1.7296	8	N.S.
B.L.	8.535	4.8780	8		10.796	4.5264	8	*
Seed	0.757	0.5276	8	N.S.	0.814	0.4218	8	N.S.
T.Input	5.055	7.6648	8	N.S.	7.746	6.3506	8	N.S.
T.Output	5.059	16.8066	8	N.S.	1.932	15.7671	8	N.S.

where H.L. stands for Human Labour/acre, B.L. stands for Bullock Labour/acre,
F.&M. " Fertilizers and Manures/acre, N.S. " Not Significant,
* " Significant at 5% level, ** " Significant at 1% level,
and (S-C) values are in rupees.

Let y_{ijk} represent the k-th observation in the (i, j)th cell, n_{ij} the number of observations in the (i,j)th cell, and m the grand mean. Further let

$$\sum_i n_{ij} = n_{.j}, \quad \sum_j n_{ij} = n_{i.}, \quad \sum_{ij} n_{ij} = \sum_j n_{.j} = \sum_i n_{i.} = n,$$

$$\sum_k y_{ijk} = Y_{ij}, \quad \sum_{jk} y_{ijk} = \sum_j Y_{ij} = Y_{i.}, \quad \sum_{ik} y_{ijk} = \sum_i Y_{ij} = Y_{.j},$$

$$\sum_{ijk} y_{ijk} = Y \quad \text{and} \quad Q_i = Y_{i.} - \sum_j \frac{Y_{.j}}{n_{.j}} n_{ij} .$$

Normal equations :

The model is $y_{ijk} = m + a_i + b_j + e_{ijk}$ (assuming no interaction) a_i being the i-th level effect of factor A, b_j the j-th level effect of factor B and e_{ijk} being the extraneous effects due to the k-th observation in the (i,j)th cell. The least square theory gives the best estimates of m, a_i and b_j by the normal equations :

$$Y_{i.} = n_{i.}m + n_{i.}a_i + \sum_j n_{ij} b_j, \qquad (i = 1,2, \ldots\ldots, p)$$

$$Y_{.j} = n_{.j}m + n_{.j}b_j + \sum_i n_{ij}a_i \qquad (j = 1,2, \ldots\ldots, q)$$

$$Y = nm + \sum_i n_{i.}a_i + \sum_j n_{.j}b_j$$

Multiplying second equation by $n_{ij}/n_{.j}$ and summing over j, we have

$$\sum_j \frac{Y_{.j}}{n_{.j}} n_{ij} = n_{i.}m + \sum_j n_{ij}b_j + \sum_j \frac{n_{ij}}{n_{.j}} \sum_i n_{ij} a_i$$

Subtracting from the first equation,

$$Q_i = (Y_{i.} - \sum_j \frac{Y_{.j}}{n_{.j}} n_{ij}) = a_i (n_{i.} - \sum_i \sum_j \frac{n_{ij}^2}{n_{.j}})$$

$$= a_i (n_{i.} - \sum_j \frac{n_{ij}^2}{n_{.j}}) - \sum_{k \neq i} a_k \sum_j \frac{n_{kj}\, n_{ij}}{n_{.j}}$$

Let $c_{ii} = (n_{i.} - \sum_j \frac{n_{ij}^2}{n_{.j}})$ and $c_{ik} = -\sum_j \frac{n_{kj}\, n_{ij}}{n_{.j}}$ $\qquad (i,k = 1,2, \ldots\ldots, p)$

Then $Q_i = \sum_k c_{ik} a_k$

To solve these equations, we note that only (p-1) of these are idependent and so for a unique solution, we impose the restriction $\sum_i a_i = 0$.

Putting $a_p = -(a_1 + a_2 + \ldots\ldots + a_{p-1})$, we have

$$Q_i = a_i P_{ii} - \sum_{k \neq i,p} a_k P_{ik} \qquad (i = 1, 2, \ldots\ldots, p-1)$$

where $P_{ii} = n_{i.} - \sum_j \frac{n_{ij}(n_{ij} - n_{pj})}{n_{.j}}$

$$P_{ik} = \sum_j \frac{n_{ij}(n_{kj} - n_{pj})}{n_{.j}}$$

Now the total s.s. viz. $\sum_{i,j,k} y^2_{ijk}$ splits up into two parts :
(i) that due to the estimates of all the constants including m and
(ii) that due to the deviation from the regression including all the constants.

The s.s. due to estimates of the constants is given by

$$Ym + \sum_i a_i Y_{i.} + \sum_j b_j Y_{.j} \quad .$$

Eliminating b_j and m, this becomes (using the normal equations)

$$\sum_i a_i Q_i + \sum_j \frac{Y^2_{.j}}{n_{.j}}$$

Again the s.s. due to the estimates of the constants on the hypothesis $a_1 = a_2 = \ldots\ldots = a_p$ is equal to $\sum_j \frac{Y^2_{.j}}{n_{.j}}$.
Hence the s.s. due to the estimates of a_i's alone i.e. s.s. adjusted for all other effects is equal to $\sum_i a_i Q_i$ based on (p-1) d.f. Let (A) denote this adjusted s.s. and A the s.s. due to the estimates of all the constants on the hypothesis $b_1 = b_2 = \ldots\ldots = b_q$ i.e. unadjusted s.s. due to the factor A. Similarly let (B) and B denote the corresponding s.s. due to the factor B.
The s.s. due to all the constants is equal to (A) + B = A + (B) from symmetry. Hence A - (A) = B-(B) = δ (say) where δ may be called the adjustment factor for non-orthogonality.

The s.s. due to the deviation from regression is

$$\sum_{i,j,k} y^2_{ijk} - (A) - B.$$

To get the interaction s.s., we use the model

$$y_{ijk} = m + a_i + b_j + h_{ij} \pm e_{ijk}$$

where h_{ij} is the interaction term of the (i,j)th cell.

The s.s. due to all the constants including h_{ij} is $\sum_{ij} \frac{Y^2_{ij}}{n_{ij}}$ with $(pq + 1)$ d.f.

Also s.s. on the assumption $h_{ij} = 0$ i.e. s.s. due to the constants m, a_i's and b_j's is (A) + B with $(p + q)$ d.f.

Hence s.s. due to interaction is $\sum_{ij} \frac{Y^2_{ij}}{n_{ij}} - (A) - B$

with $(p-1)(q-1)$ d.f.

Therefore s.s. due to deviation from regression is

$$\sum_{ijk} y^2_{ijk} - \sum_{ij} \frac{Y^2_{ij}}{n_{ij}} \text{ based on } (n-pq-1) \text{ d.f.}$$

In particular when A has only two classes say a_1 and a_2 we have

$$a_1 = -a_2 = \frac{Q_1}{2\sum_j \frac{n_{1j} n_{2j}}{n_{.j}}} = \frac{Q_1}{2P} \quad \text{(say).}$$

The adjusted s.s. due to A is given by $(A) = \frac{Q^2_1}{P}$

and variance of estimate of $(a_1 - a_2) = \frac{\sigma^2}{P}$

Let estimate of $\sigma^2 = \hat{\sigma}^2 = s^2$

Then $F_{(1,n)} = \frac{Q^2_1}{P} / s^2$; or using t - test, if $\bar{x}$ and $\bar{y}$ be the overall means of the two methods, then

$$\bar{d} = (\bar{y} - \bar{x}) = \frac{2\,Q_1}{2\,P} = \frac{Q_1}{P}$$

$$V(\bar{d}) = V(\bar{y} - \bar{x}) = \frac{s^2}{P}$$

$$S.E.(\bar{d}) = \frac{s}{\sqrt{P}}$$

$$t_n = \frac{\bar{d}}{S.E.(\bar{d})} = \frac{Q_1\sqrt{P}}{P\,s} = \frac{Q_1}{s\sqrt{P}}$$

If on the other hand, the interaction s.s. is significant, the s.s. due to the methods changes and is calculated by using appropriate weights as follows:

The following table gives the means of the groups within a village for the methods. The numbers are the number of observations on which the corresponding mean is based.

Cost Accounting		Survey	
Group Mean	No. of observations	Group Mean	No. of observations
$\bar{x}_1$	n_1	$\bar{y}_1$	m_1
$\bar{x}_2$	n_2	$\bar{y}_2$	m_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\bar{x}_k$	n_k	$\bar{y}_k$	m_k

Let $\bar{x}$ and $\bar{y}$ be the overall means for the two methods and s^2 the overall mean square within cells with n d.f.

$$\text{Then } V(\bar{x}) = \frac{s^2}{k^2}\left(\frac{1}{n_1}+\frac{1}{n_2}+\cdots\cdots+\frac{1}{n_k}\right) = \frac{s^2}{N} \text{ (say)}$$

$$V(\bar{y}) = \frac{s^2}{k^2}\left(\frac{1}{m_1}+\frac{1}{m_2}+\cdots\cdots+\frac{1}{m_k}\right) = \frac{s^2}{M} \text{ (say)}$$

$$\text{where } N = k^2 \Big/ \left(\frac{1}{n_1}+\frac{1}{n_2}+\cdots\cdots+\frac{1}{n_k}\right)$$

$$\text{and } M = k^2 \Big/ \left(\frac{1}{m_1}+\frac{1}{m_2}+\cdots\cdots+\frac{1}{m_k}\right)$$

$$\text{Let } \bar{d} = \bar{y} - \bar{x}.$$

$$\text{Then } V(\bar{d}) = V(\bar{y}-\bar{x}) = V(\bar{y}) + V(\bar{x}) = s^2\left(\frac{1}{N}+\frac{1}{M}\right)$$

$$\text{and } t_{(n\,d.f.)} = \bar{d} \Big/ s\sqrt{\frac{1}{N}+\frac{1}{M}}$$

Also s.s. due to the methods (adjusted for the disproportionateness in cell frequencies) is given by

$$(A) = N\bar{x}^2 + M\bar{y}^2 - \frac{(N\bar{x}+M\bar{y})^2}{N+M}$$

$$F_{(1,n)} = \left\{N\bar{x}^2 + M\bar{y}^2 - \frac{(N\bar{x}+M\bar{y})^2}{N+M}\right\} \Big/ s^2 .$$

To see the equivalance of t and F, we have

$$F = \left\{N^2\bar{x}^2 + M^2\bar{y}^2 + NM(\bar{x}^2+\bar{y}^2) - N^2\bar{x}^2 - M^2\bar{y}^2 - 2NM\bar{x}\bar{y}\right\} \Big/ s^2(N+M)$$

$$= \frac{NM(\bar{y}-\bar{x})^2}{s^2(N+M)} = \frac{(\bar{y}-\bar{x})^2}{s^2\left(\frac{N+M}{NM}\right)}$$

$$= \bar{d}^2 \Big/ s^2\left(\frac{1}{N}+\frac{1}{M}\right) = t^2.$$

Hence the two approaches are equivalent.

As t-test has been employed in the comparison between the two methods based on independent villages, the same will be used in the comparison based on common villages. The procedure followed is that first we test the interaction s.s. in each case. In case it is not significant we use $t = \frac{Q_1}{s\sqrt{P}}$, if it is significant, we have $t = \frac{\bar{d}}{s\sqrt{\frac{1}{N}+\frac{1}{M}}}$ where all these quantities have already been defined. Hence, after illustration by dealing with a few cases in details, appropriate form of t-test has been used after testing the interaction s.s. in each case.

Some examples:

Wheat, Human labour

(i) Village Gagore

Cost Accounting		Survey		Total	
No. of holdings	Sum of the obsns. in each cell.	No. of holdings.	Sum of the obsns. in each cell.	No.	Sum.
2	85.84	1	48.01	3	133.85
2	71.90	3	147.39	5	219.29
2	54.26	4	140.28	6	194.54
2	68.26	4	162.92	6	231.18
2	51.64	4	182.84	6	234.48
Total 10	331.90	16	681.44	26	1013.34

Now $Q_1 = X_{1.} - \sum_j \frac{X_{.j}}{n_{.j}} n_{1j}$ = 331.90 − 89.23 − 87.72 − 64.85 − 77.06 − 78.16 = −66.11

$P = \sum_j \frac{n_{1j} n_{2j}}{n_{.j}}$ = 2/3 + 6/5 + 4/3 + 4/3 + 4/3 = 88/15

The adjusted s.s. due to methods (A) = Q_1^2/P = $(-66.11)^2 \frac{15}{88}$

= 744.977 with 1 d.f.

S.S. due to groups B = 5971.941 + 9617.621 + 6307.635 + 8907.365 + 9163.478 − 39494.537 = 473.503 with 4 d.f.

Between cells s.s. = Y_{ij}^2/n_{ij} = 3684.253 + 2584.805 + 1472.074 + 2329.714 + 1333.345 + 8804.960 + 7241.271 + 4919.620 + 6635.732 + 8362.188 - 39494.537 = 1373.425 with 9 d.f.

Interaction s.s. = 1373.425 - 744.977 -473.503 = 154.945 with 4 d.f.

Total s.s. = 11881.250 + 31087.550 - 39494.536 = 3472.263

Error s.s. = 2100.838 with 16 d.f.

Analysis of Variance Table

Source of Variation	D.F.	S.S.	M.S.	F
Methods	1	744.977	744.977	6.605 *
Groups	4	473.503	118.376	
Interaction	4	154.945	38.736	N.S.
Error	16	2100.838	131.302	
Pooled Error	20	2255.783	112.789	

Note that the negative sign of Q shows that the Survey method gives significantly greater values.

$\bar{d}$ = (S-C) = 11.269 V($\bar{d}$) = 19.2989

S.E.($\bar{d}$)=4.393 $t_{20 d.f.}$ =2.565 *

(ii) Wheat Human Labour

Village Bajhera Kalan

Q_1 = -113.63,

P = 19/3 , (A) = 2039.060 with 1 d.f.

A.V.Table

Source	D.F.	S.S.	M.S.	F
Methods	1	2039.060	2039.060	6.106 *
Groups	4	403.498	100.875	
Interaction	4	495.833	123.958	N.S.
Error	18	6850.422	380.579	
Pooled Error	22	7346.255	333.920	

$\bar{d}$ =17.942 V($\bar{d}$) = 52.7244

S.E.($\bar{d}$) = 7.261 $t_{22 d.f.}$ = 2.471 *

Wheat. Bullock Labour

(iii) Village Jasora

A.V.Table

Source	D.F.	S.S.	M.S.	F
Methods	1	13221.872	13221.872	
Groups	4	97610.264		
Interaction	4	42116.264	10529.066	3.199 *
Error	15	--	3291.749	

Here the interaction is significant.

$\therefore$ Weighted $\bar{\bar{d}} = 67.700$ $V(\bar{d}) = 713.2122$

$S.E.(\bar{\bar{d}}) = 26.706$ $t_{(15\ d.f.)} = 2.535$ *

Having dealt with a few examples un details to illustrate the approaches suitable for the two cases viz. (i) when the interaction is absent and (ii) when it is present, in the following tables are given the results obtained for all the villages in the above manner.

Table 5.2

A village-wise comparison of the two methods of collecting data

Wheat.

Village	(S-C)	S.E.	D.F.	t	
Human Labour:					
Datiyana	8.019	3.129	18	2.563	*
Gagore	11.269	4.393	20	2.565	*
Khandraoli	14.129	4.294	22	3.290	**
L.Rasulpur	-1.096	3.743	22	0.266	N.S.
B.Kalan	17.942	7.261	22	2.471	*
Jasora	9.094	5.266	19	1.727	N.S.
Bullock Labour:					
Datiyana	60.758	16.241	18	3.741	**
Gagore	86.028	16.478	20	5.221	**
Khandraoli	105.376	24.701	22	4.266	**
L.Rasulpur	94.807	19.054	22	4.976	**
B.Kalan	34.948	24.941	22	1.401	N.S.
Jasora	67.700	26.706	18	2.535	**
Seed:					
Datiyana	1.121	0.4612	18	2.431	*
Gagore	3.750	0.4089	20	9.171	**
Khandraoli	2.821	0.4568	22	4.174	**
L.Rasulpur	0.472	0.6079	22	0.776	N.S.
B.Kalan	0.341	1.0969	22	0.311	N.S.
Jasora	3.002	1.1530	19	2.604	*

(continued on next page)

Wheat.

Total Input:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	84.437	19.644	18	4.298	**
Gagore	108.224	20.173	16	5.365	**
Khandraoli	160.953	37.022	22	4.347	**
L.Rasulpur	93.935	22.869	22	4.107	**
B.Kalan	44.547	33.048	22	1.348	N.S.
Jasora	97.224	30.675	15	3.170	**

Total Output:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	89.308	22.674	18	3.939	**
Gagore	6.598	23.534	20	0.280	N.S.
Khandraoli	59.837	22.194	22	2.696	*
L.Rasulpur	24.058	27.413	22	0.828	N.S.
B.Kalan	-18.725	15.722	22	1.191	N.S.
Jasora	-14.748	19.285	19	0.765	N.S.

Table 5.3

A comparison of the two methods of collecting data (village-wise)

Sugarcane (planted).

Village	(S-C)	S.E.	D.F.	t	
Human Labour:					
Datiyana	53.960	22.299	21	2.420	*
Gagore	50.235	13.895	18	3.615	**
Sikri	26.872	6.737	15	3.980	**
Khandraoli	45.962	12.818	16	3.586	**
L.Rasulpur	1.037	14.012	18	0.074	N.S.
B.Kalan	5.726	27.354	8	0.209	N.S.
Jasora	46.048	16.582	16	2.777	*
Bullock Labour:					
Datiyana	62.745	17.475	21	3.591	**
Gagore	200.094	69.653	18	2.873	*
Sikri	112.090	13.058	15	8.584	**
Khandraoli	-20.953	12.147	16	1.725	N.S.
L.Rasulpur	17.394	15.069	18	1.156	N.S.
B.Kalan	-23.285	29.139	8	0.799	N.S.
Jasora	76.693	21.273	16	3.605	**
Seed:					
Datiyana	-5.935	2.938	21	2.020	N.S.
Gagore	25.498	4.698	18	5.428	**
Sikri	-4.565	6.516	19	0.700	N.S.
Khandraoli	-0.478	3.398	13	0.141	N.S.
L.Rasulpur	-17.083	8.465	18	2.018	N.S.
B.Kalan	-4.266	8.756	8	0.487	N.S.
Jasora	-10.905	8.124	16	1.342	N.S.

(contd. on next page)

Sugarcane (planted).

Fertilizers and Manures:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	-1.881	9.455	19	0.199	N.S.
Sikri	-13.317	6.933	19	1.921	N.S.
Khandraoli	1.130	9.504	12	0.119	N.S.
L.Rasulpur	-21.498	11.300	13	1.903	N.S.
B.Kalan	-6.481	10.144	8	0.639	N.S.
Jasora	7.996	12.159	15	0.658	N.S.

Total Input:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	132.737	42.224	21	3.144	**
Gagore	320.001	90.990	18	3.517	**
Sikri	123.008	27.369	15	4.494	**
Khandraoli	10.973	23.472	16	0.467	N.S.
L.Rasulpur	-7.035	39.432	18	0.178	N.S.
B.Kalan	-7.108	78.006	8	0.091	N.S.
Jasora	106.144	37.610	16	2.823	*

Total Output:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	151.605	97.162	21	1.560	N.S.
Gagore	221.224	90.801	18	2.436	*
Sikri	115.831	31.536	19	3.673	**
Khandraoli	263.852	111.999	16	2.356	*
L.Rasulpur	126.190	100.865	18	1.251	N.S.
B.Kalan	71.011	85.885	8	0.827	N.S.
Jasora	61.999	116.490	16	0.532	N.S.

Table 5.4

A comparison of the two methods of collecting data (village-wise)

<u>Gram</u>.

Village	(S-C)	S.E.	D.F.	t	
<u>Human Labour</u>:					
Datiyana	3.192	2.800	19	1.140	N.S.
Gagoro	5.115	1.630	23	3.137	**
Sikri	-6.062	1.633	13	3.712	**
Khandraoli	1.841	3.380	10	0.545	N.S.
L.Rasulpur	-10.476	3.035	12	3.451	**
B.Kalan	6.693	7.355	5	0.910	N.S.
Jasora	-1.294	7.180	8	0.180	N.S.
<u>Bullock Labour</u>:					
Datiyana	56.029	16.967	20	3.302	**
Gagore	12.998	7.127	23	1.824	N.S.
Sikri	36.559	9.838	13	3.716	**
Khandraoli	38.216	10.575	10	3.614	**
L.Rasulpur	17.974	9.429	16	1.906	N.S.
B.Kalan	-4,880	12.663	4	0.385	N.S.
Jasora	-5.053	18.706	8	0.270	N.S.
<u>Seed</u>:					
Datiyana	1.370	0.4744	20	2.902	**
Gagore	0.135	0.2053	23	0.658	N.S.
Sikri	-1.862	0.4195	13	4.440	**
Khandraoli	-1.300	0.5605	10	2.319	*
L.Rasulpur	1.175	0.6358	16	1.847	N.S.
B.Kalan	-1.020	1.2459	5	0.819	N.S.
Jasora	2.472	2.0720	8	1.193	N.S.

(contd. on next page)

Gram.

Total Input:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	41.340	19.315	20	2.140	*
Gagore	19.381	7.906	23	2.451	*
Sikri	31.466	11.770	13	2.673	*
Khandraoli	40.031	13.781	10	2.905	*
L.Rasulpur	0.204	15.080	16	0.014	N.S.
B.Kalan	15.948	24.257	5	0.657	N.S.
Jasora	-4.426	18.110	8	0.244	N.S.

Total Output:

Village	(S-C)	S.E.	D.F.	t	
Datiyana	65.099	28.350	20	2.296	*
Gagore	00.081	8.324	23	0.010	N.S.
Sikri	18.156	12.406	13	1.463	N.S.
Khandraoli	-2.659	22.017	10	0.121	N.S.
L.Rasulpur	-25.540	18.738	16	1.363	N.S.
B.Kalan	-8.400	12.183	4	0.689	N.S.
Jasora	17.080	20.968	8	0.815	N.S.

5.4 Having dealt with village-wise comparison of the two methods based on common villages, we now combine the results over the villages in each stratum. For this purpose, as the villages have been selected with probability proportional to their areas, equal weights for the villages within a stratum have been used. Thus we have the following tables giving zonal-wise comparison .

Table 5.5
A zonal-wise comparison between the two methods (common villages).

Wheat.

Zone	(S-C)	S.E.	D.F.	t	
Human Labour:					
I.	9.644	2.697	38	3.576	**
II	14.129	4.292	22	3.292	**
III	-1.096	3.743	22	0.293	N.S.
IV	13.516	4.485	41	3.014	**
Bullock Labour:					
I	73.390	11.568	38	6.344	**
II	105.376	24.701	22	4.266	**
III	94.807	19.054	22	4.976	**
IV	51.324	18.271	37	2.809	**
Seed:					
I	2.437	0.3082	38	7.907	**
II	2.821	0.4568	22	6.175	**
III	0.472	0.6079	22	0.776	N.S.
IV	1.672	0.7940	41	2.106	*
Total Input:					
I	96.329	14.079	38	6.842	**
II	160.953	37.022	22	4.347	**
III	93.935	22.869	22	4.107	**
IV	70.886	17.223	37	4.116	**
Total Output:					
I	47.953	16.340	38	2.935	**
II	59.837	22.194	22	2.696	*
III	24.058	27.413	22	0.878	N.S.
IV	-16.737	12.441	41	1.345	N.S.

Table 5.6

A comparison (zonal-wise) between the two methods (common villages)

Sugarcane (planted).

Human Labour:

Zone	(S-C)	S.E.	D.F.	t	
I	52.098	13.137	39	3.966	**
II	36.417	7.240	31	5.030	**
III	1.037	14.018	18	0.740	N.S.
IV	25.887	15.994	24	1.619	N.S.

Bullock Labour:

Zone	(S-C)	S.E.	D.F.	t	
I	131.420	35.906	39	3.660	**
II	45.569	8.917	31	5.110	**
III	17.394	15.069	18	1.154	N.S.
IV	26.704	18.039	24	1.480	N.S.

Seed:

Zone	(S-C)	S.E.	D.F.	t	
I	9.782	2.770	39	3.531	**
II	-2.521	3.674	32	0.686	N.S.
III	-17.083	8.465	18	2.018	N.S.
IV	-7.686	6.972	24	1.270	N.S.

Fertilizers and Manures:

Zone	(S-C)	S.E.	D.F.	t	
I	-1.881	9.455	19	0.199	N.S.
II	-6.094	5.882	31	1.036	N.S.
III	-21.498	11.300	13	1.903	N.S.
IV	0.758	7.917	23	0.096	N.S.

Total Input:

Zone	(S-C)	S.E.	D.F.	t	
I	226.369	50.155	39	4.513	**
II	66.991	18.028	31	3.776	**
III	-7.036	39.432	18	0.178	N.S.
IV	48.518	42.499	24	1.143	N.S.

(continued on next page)

Sugarcane (planted).

Total Output:

Zone	(S-C)	S.E.	D.F.	t	
I	186.415	66.493	39	2.804	**
II	189.842	58.177	35	3.263	**
III	126.190	100.865	18	1.251	N.S.
IV	66.505	72.364	24	0.919	N.S.

Table 5.7

A zonal-wise comparison between the two methods (common villages).

Gram.

Zone	(S-C)	S.E.	D.F.	t	
Human Labour:					
I	4.154	1.619	42	2.565	*
II	2.111	1.877	23	1.124	N.S.
III	-10.476	3.075	12	3.451	**
IV	2.700	5.138	13	0.525	N.S.
Bullock Labour:					
I	34.514	9.202	43	3.751	**
II	37.369	7.222	23	5.177	**
III	17.974	9.429	16	1.906	N.S.
IV	-4.967	9.330	12	0.532	N.S.
Seed:					
I	-0.618	0.2583	43	2.391	*
II	-1.581	0.3501	23	4.516	**
III	1.175	0.6358	16	1.847	N.S.
IV	0.726	0.3010	13	2.414	*
Total Input:					
I	30.361	10.435	43	2.910	**
II	35.749	9.062	23	3.945	**
III	0.204	15.080	16	0.013	N.S.
IV	5.760	1.514	13	3.805	**
Total Output:					
I	32.590	14.773	43	2.206	*
II	7.749	12.636	23	0.613	N.S.
III	-25.540	18.738	16	1.363	N.S.
IV	4.340	12.125	12	0.358	N.S.

5.5 We now have to sum up the results (obtained above) over tho different zones for any particular factor. To obtain the overall difference of means of Survey and Cost Accounting methods we may use tho area of the zone as the weight for the particular mean. Thus if Ai be the area of the i-th zone and $\bar{d}_i$ the difference of means for a factor in that zone, the overall difference of means for that factor is given by

$$\bar{d} = \sum_i \frac{A_i \bar{d}_i}{\sum A_i} = \sum_i \lambda_i \bar{d}_i \quad \text{where } \lambda_i = \frac{A_i}{\sum_i A_i}$$

and $$V(\bar{d}) = \sum_i \frac{A_i^2 V(\bar{d}_i)}{(\sum A_i)^2} = \sum_i \lambda_i^2 V(\bar{d}_i)$$ as each $\bar{d}_i$ is independent.

Thus the appropriate S.E. is obtained. We test $\bar{d}$ against this S.E. using $t = \bar{d}/S.E.(\bar{d})$ with appropriate d.f.

The second method of combining and the results is to used equal weights for the various zones. The former method would be justified if we wish to take account of the variation in the cultivating population of each village for reasons explained earlier. (The area figures are used for weights as figures for cultivating population are not available). The latter approach would be justified if we treat the present investigators as a random sample of investigators and give equal weight to each investigator.

Taking one case in detail, we have

Table No. 5.

<u>Wheat</u> Human Labour

Zone	(S - C)	Variance	=	
I	9.644	7.2714	0.229564	0.052700
II	14.129	18.4383	0.185207	0.034302
III	-1.096	14.0066	0.256832	0.065963
IV	13.096	20.1123	0.328397	0.107845

Using proportional weights, the overall difference of means, d, is given by $\bar{d} = \sum_i \frac{A_i \bar{d}_i}{\sum_i A_i} = 8.9885$ $V(\bar{d}) = \sum_i \frac{A_i^2}{(\sum A_i)^2} V(\bar{d}_i) = 4.1086$

S.E.$(\bar{d})$ = 2.0270 $t_{(12)} = 4.434$ **

If on the other hand, we use equal weights, we have

$(\bar{d}) = 9.0488 = \frac{1}{n}\sum_i \bar{d}_i$ $\qquad V(\bar{d}) = \frac{1}{n^2}\sum_i V(\bar{d}_i) = 3.7393$

$S.E.(\bar{d}) = 1.9337$ $\qquad t_{(n-3)} = 4.670$ **

5.6 We are now in a position to set out the two types of comparisons viz. that based on independent villages and that based on common villages.

Table 5.8

Weights proportional to areas.

Factor	Independent Villages (S-C)	S.E.	D.F.	t		Common Villages (S-C)	S.E.	D.F.	t	
Wheat.										
H.L.	2.509	3.4464	8	0.728	N.S.	8.989	2.0270	123	4.434	**
B.L.	70.768	13.1975	8	5.362	**	77.568	9.3771	119	8.272	**
Seed	-0.680	0.7618	8	0.893	N.S.	1.752	0.3976	123	4.407	**
T.Input	77.424	16.3098	8	4.747	**	99.328	11.1331	115	8.922	**
T.Output	-14.456	16.8363	8	0.859	N.S.	22.773	9.8605	123	2.310	**
Sugarcane (planted).										
H.L.	-0.456	11.1292	8	0.041	N.S.	27.472	7.1715	112	3.831	**
B.L.	48.924	16.2876	8	2.967	*	51.846	1.9883	112	4.718	**
Seed	0.402	5.6991	8	0.071	N.S.	-5.125	3.0727	113	1.668	N.S.
F.&M.	-20.661	7.0352	8	2.937	*	-6.833	4.5913	86	1.488	N.S.
T.Input	26.529	25.6307	8	1.035	N.S.	78.828	21.1710	112	3.730	**
T.Output	-30.440	62.6895	8	0.486	N.S.	132.200	39.8112	116	3.321	**
Gram.										
H.L.	-3.103	1.869	8	1.660	N.S.	-0.459	1.9271	90	0.238	N.S.
B.L.	8.535	4.8780	8	1.750	N.S.	17.833	4.6372	94	3.846	**
Seed	-0.757	0.5276	8	0.435	N.S.	0.117	0.3635	95	0.322	N.S.
T.Input	5.055	7.1643	8	0.706	N.S.	15.535	6.9470	95	2.236	*
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	94	0.505	N.S.

Table 5.9

Equal weights.

Factor	Independent Villages (S-C)	S.E.	D.F.	t		Common Villages (S-C)	S.E.	D.F.	t	
Wheat:										
H.L.	1.877	3.8708	8	0.485	N.S.	9.049	1.9337	123	4.670	**
B.L.	73.658	4.2441	8	17.355	**	81.224	8.4896	119	8.559	**
Seed	-0.452	0.7245	8	0.624	N.S.	1.851	0.3878	123	4.772	**
T.Input	79.794	18.3306	8	4.353	**	105.526	12.2179	115	8.637	**
T.Output	-14.310	15.8328	8	0.904	N.S.	28.779	10.2037	123	2.820	**
Sugarcane (planted):										
H.L.	4.179	10.5706	8	0.395	N.S.	28.860	6.5055	112	4.436	**
B.L.	55.429	14.8000	8	3.745	**	54.772	10.9477	112	5.003	**
Seed	2.400	5.0275	8	0.477	N.S.	-4.352	2.8340	113	1.536	N.S.
F.&M.	-16.419	6.0880	8	2.697	*	-7.179	4.4325	86	1.620	N.S.
T.Input	41.747	23.0187	8	1.814	N.S.	83.961	19.7900	112	4.243	**
T.Output	-31.937	54.8575	8	0.582	N.S.	142.238	38.0922	116	3.734	**
Gram:										
H.L.	-3.037	1.7296	8	1.756	N.S.	-0.378	1.6155	90	0.234	N.S.
B.L.	10.796	4.5264	8	2.385	*	21.228	4.4214	94	4.801	**
Seed	-0.814	0.4218	8	0.193		-0.059	0.3062	95	0.192	N.S.
T.Input	7.746	6.3506	8	1.220	N.S.	18.019	6.3615	95	2.832	**
T.Output	-1.932	15.7671	8	0.123	N.S.	4.785	7.3995	94	0.647	N.S.

From the set of tables, just given, we see that proportional weights and equal weights give similar results. The difference in the two types of comparisons is that the one based on common villages gives more significant results than the one based on independent villages, but there is no contradiction in the two types of comparisons i.e., there is no factor for which one type of comparison gives significantly different results in favour of (say) the Survey method and the same factor in the other type of comparison gives a significantly different result in favour of the Cost Accounting method. As a matter of fact, this similarity of results might have been expected and we have much more degrees of freedom in the common villages. We can now combine the results over the two types of comparisons to obtain an overall picture of the difference for the various factors in the three crops separately. This combination is done by using the inverse of the variance of a difference of means as the weight for that difference. Thus, if w_i be the inverse of variance of difference of means $\bar{d}_i$, the combined difference of means is given by

$$\bar{d} = \sum_i \frac{w_i \bar{d}_i}{\sum_i w_i}$$

The variance of $\bar{d}$ is given by

$$V(\bar{d}) = \sum_i \frac{w_i^2}{(\sum_i w_i)^2} V(\bar{d}_i)$$ as each $\bar{d}_i$ is independent.

Thus we obtain the appropriate standard error for testing $\bar{d}$.

Combining the previous results over the two types of comparisons, we have the following tables.

Table 5.10

	Proportional Weights					Equal Weights				
Wheat:										
Factor	(S-C)	S.E.	D.F.	t		(S-C)	S.E.	D.F.	t	
H.L.	7.3236	1.7472	131	4.192	**	7.6166	1.7299	131	4.403	**
B.L.	75.2867	7.6441	127	9.849	**	78.8982	7.8975	127	9.990	**
Seed	1.2313	0.3525	131	3.493	**	1.3381	0.3419	131	3.914	**
T.Input	92.3658	9.1926	123	10.048	**	97.6107	10.1665	123	9.601	**
T.Output	13.2648	8.5086	131	1.559	N.S.	16.1343	8.5769	131	1.881	N.S.
Sugarcane (planted):										
H.L.	19.8779	6.0283	120	3.198	**	22.0800	5.5403	120	3.985	**
B.L.	50.9321	9.1091	120	5.591	**	55.0044	8.8016	120	6.250	**
Seed	-3.8803	2.7046	121	1.435	N.S.	-2.7259	2.4688	121	1.103	N.S.
F.&M.	-10.9633	3.8450	94	2.851	**	-10.3801	3.5834	94	2.897	**
T.Input	52.3877	15.0476	120	3.481	**	66.0201	15.0064	120	4.399	**
T.Output	85.4601	33.6071	124	2.543	*	85.5778	31.2887	124	2.735	**
Gram:										
H.L.	-1.8215	1.3417	98	1.358	N.S.	-1.6169	1.1806	98	1.370	N.S.
B.L.	13.4191	3.3609	102	3.773	**	16.1344	3.1629	102	5.105	**
Seed	-0.1643	0.2993	103	0.549	N.S.	-0.3195	0.2477	103	1.290	N.S.
T.Input	10.4765	4.9966	103	2.097	*	12.8738	4.4944	103	2.864	**
T.Output	2.3193	6.8360	102	0.339	N.S.	3.8726	6.6316	102	0.539	N.S.

5.7 From the foregoing tables, we see that there is no difference in the results obtained whether we use proportional weights or equal weights. Thus we may use either of the two final tables for basing our conclusion upon. Another fact that emerges out of these tables is that the factors Bullock Labour/acre which is the major item of input as also the Total Input/acre show a significant difference in all the three crops. The magnitude of the difference expressed as percentage increase over the Cost Accounting values, in the case of Bullock Labour/acre being of the order of 70% for Wheat, 55% for Sugarcane (planted) and 40% for the Gram crop. This percentage excess in the case of Total Input/acre for the above three crops respectively is of the order of 50%, 15% and 10%. Regarding the factor Human Labour/acre we see that it shows significant difference in the case of Wheat and Sugarcane (planted) but no significant difference in Gram; the percentage excess being of the order of 20%, 20% and 10%. The fact Seed/acre is significant only in Wheat, the excess being of the order of 10%. The factor Fertilizers and Manures/acre which was applied to Sugarcane crop only shows a significant difference (in the favour of Cost Accounting method) , the excess being of the order of 25%. Finally taking the case of Total Output/acre, we see that it shows a significant difference in Sugarcane, the excess being of the order 15%. Thus we see that under the Survey method the expenses have usually been reported greater and that the excess does not follow any regular pattern. Thus the possibility of adjustment in the Survey figures by some predetermined process is ruled out. Hence it would be advisable to employ the Cost Accounting method for collecting the primary data despite the cost involved.

6. Estimation of Sample Size for given precision

6.1 The third part of the present study concerns the sampling design viz. the estimation of the sample size for a given precision. In the present survey there are 20 villages with 10 holdings selected from each village. The purpose of this section is to estimate the total number of villages to be selected as also the number of holdings in each selected village so that the percentage standard error of the estimate (of the population mean) is some pre-assigned quantity say p (The usual values of p being 0.05 and 0.01).

6.2 Let y_{tij} be the j-th holding of i-th village in the t-th stratum. As the villages have been selected with probability proportional to the area, let P_{ti} be the probability of selection assigned to the i-th village in the t-th stratum (henceforth to be denoted by (t,i) th village). If A_{ti} be the area of the (t,i)th village and A_t the area of the t-th stratum,

we have $P_{ti} = A_{ti}/A_t$

Let R_{ti} and R_t denote the corresponding (cultivating) population figures.

Let $$z_{tij} = \frac{R_{ti}}{R_t} \; \frac{1}{P_{ti}} \; y_{tij}$$

As the actual values of R_{ti} and R_t are not available, assuming $R_{ti} \propto A_{ti}$, we estimate (R_{ti}/R_t) by (A_{ti}/A_t).

Then $$z_{tij} = \frac{A_{ti}}{A_t} \; \frac{1}{P_{ti}} \; y_{tij} = y_{tij}$$

Let K be the number of strata in the population,

N_t be the number of village in the t-th stratum,

M_{ti} be the number of holdings in the (t,i)th village,

n_t be the number of villages selected from the t-th stratum,

m_{ti} be the number of holdings selected from the (t,i)th village

Further, let

$\bar{z}_{...}$ be the population mean,

$\bar{z}_{t..}$ be the mean of the t-th stratum,

$\bar{z}_{ti.}$	be the mean of the (t,i) village
$\bar{z}$	be the sample estimate of the population mean
$\bar{z}_{t(nt)}$	be the sample estimate of the mean of the t-th stratum
$\bar{z}_{ti(mti)}$	be the sample estimate of the mean of the (t,i) i-th village,

Then $\bar{z}_{ti}(mti) = \frac{1}{mti} \cdot \sum_{j}^{m_{ti}} z_{tij}$

gives an unbiased estimate of the (t,i)th village mean.

Similarly the unbiased estimate of the t-th stratum mean is given by

$$\bar{z}_{t(n_t)} = \frac{1}{n_t} \sum_{i}^{n_t} \bar{z}_{ti(m_{ti})}$$

Hence the unbiased estimate of the population mean is given by

$$\bar{z} = \sum_{t=1}^{k} \lambda_t \, \bar{z}_{t(n_t)}$$

6.2.1 where $\lambda_t = A_t / \sum_{t=1}^{k} A_t$.

and its variance is given by

6.2.2 $$V(\bar{z}) = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \left[\sigma_{tb}^2 + \sum_{i=1}^{N_t} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) S_{ti}^2 \right]$$

where $\sigma_{tb}^2 = \sum_{i=1}^{N_t} P_{ti} \, (\bar{z}_{ti.} - \bar{z}_{t..})^2$

and $S_{ti}^2 = \frac{1}{M_{ti}-1} \sum_{j=1}^{M_{ti}} (z_{tij} - \bar{z}_{ti.})^2$

It may be remarked here that as we are interested in the sample size of holdings within a village and not within a group of holdings (which is arbitrary) of a village, we consider the ten holdings in each village to be a random sample out of all the holdings of the village.

The estimate of ($\bar{z}$) is given by :

6.2.3 $$\text{Est. } V(\bar{z}) = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} s_{tb}^2$$

where $s_{tb}^2 = \frac{1}{n_t - 1} \sum_{i}^{n_t} \left\{ z_{ti(m_{ti})} - \bar{z}_{t(n_t)} \right\}^2$

Also
6.2.4 $$\begin{cases} \text{Est. } \sigma_{tb}^2 = \hat{\sigma}_{tb}^2 = s_{tb}^2 - \frac{1}{n_t} \sum_{i}^{n_t} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) s_{ti}^2 \\ \text{Est. } S_{ti}^2 = s_{ti}^2 \end{cases}$$ and

6.2.5 where $s_{ti}^2 = \frac{1}{m_{ti}-1} \sum_{j}^{m_{ti}} \left\{ z_{tij} - \bar{z}_{ti(m_{ti})} \right\}^2$

Using (6.2.4) $\text{Est } V(\bar{z}) = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \left[\hat{\sigma}_{tb}^2 + \frac{1}{n_t} \sum_{i}^{n_t} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) s_{ti}^2 \right]$

or neglecting $1/M_{ti}$ which is small compred to 1/mti we have

$$\text{Est. } V(\bar{z}) = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \left[\hat{\sigma}_{tb}^2 + \frac{1}{n_t} \sum_{i}^{n_t} \frac{s_{ti}^2}{m_{ti}} \right]$$

Evaluating $\hat{\sigma}^2_{tb}$ and s^2_{ti} from the sample, we substitute in (6.2.6) which is a function of n_t and m_{ti}. For a given precision, fixing m_{ti}, we get the corresponding value of n_t.

In actual practice, we put $m_{ti} = m$ i.e. we select the same number of holdings from each village. Then putting $= B_t$, and $(1/n_t)\sum_i^{n_t} s^2_{ti} = W_t$, we have

$$\text{Est.}V(\bar{\bar{z}}) = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}(B_t + W_t/m) \quad \text{where } B_t \text{ is the} \qquad (6.2.7)$$

between villages means square and W_t the within villages (pooled) mean square for the t-th stratum. From (6.2.7) we can determine the number of villages with a given precision for each stratum separately. Adding these, we get the overall sample size. But a bett r approach, and the one which is usually adopted in such surveys, is that we determine the overall sample size (n,m) and then allocate n over the different strata according to plan say e.g. equal or proportional allocation etc. Thus assuming $(B_t + W_t/m)$ to be uniform over the strata, we have, denoting $(B_t + W_t/m)$ by $(B + W/m)$,

$$\begin{aligned}
\text{Est.} \quad V(\bar{\bar{z}}) &= (B + W/m)\sum_{t=1}^{k}\frac{\lambda_t^2}{n_t} \\
&= (B + W/m)\sum_{t=1}^{k}\frac{1}{n_t}\left(\frac{M_t}{\Sigma M_t}\right)^2 \\
&= (B + W/m)\sum_{t=1}^{k}\frac{1}{n_t}\left(\frac{n_t}{\Sigma n_t}\right)^2 \quad \text{putting } n_t \propto M_t. \\
&= (B + W/m)\sum_{t=1}^{k}\frac{n_t}{(\sum_{t=1}^{k} n_t)^2} \\
&= 1/n\,(B + W/m).
\end{aligned}$$

6.2.8 i.e. $\quad \text{Est. } V(\bar{z}) = \frac{B}{n} + \frac{W}{mn}$

6.2.9 Thus we have

$$(B/n + W/mn) = V_0$$

where V_0 is the given variance, B and W being known from the sample, fixing m,n can be determined. In this way we shall have a table of values of m and n. Now the most important consideration in any survey, is however, the cost involved. The simplest possible cost function that we may consider for our survey is of the type

6.2.10 $C_0 = c_1 n + c_2 mn$

where c_1 is the cost per village and c_2 is the cost per holding, being taken to be the same from village to village. The constant cost factor i.e. overhead cost or the cost of maintainace at the headquarters is omitted for calculations.

Now for least cost, we have that

$$(B/n + W/mn) + \mu(c_1 n + c_2 mn) \text{ should be minimum.}$$

Minimisation gives

$$n = \sqrt{B/\mu c_1} \qquad mn = \sqrt{W/\mu c_2}$$

$$\text{i.e.} \quad m = \sqrt{\frac{W c_1}{B c_2}}$$

Substituting for m and n in (6.2.9), we have

$$(\sqrt{B\mu c_1} + \sqrt{W\mu c_2}) = V_0$$

Therefore

$$\sqrt{\mu} = \frac{V_0}{\sqrt{Bc_1} + \sqrt{Wc_2}}$$

6.2.11

$$\text{i.e.} \quad n = \frac{\sqrt{B}\,(\sqrt{Bc_1} + \sqrt{Wc_2})}{V_0\sqrt{c_1}} = \frac{1}{V_0}\left(B + \sqrt{\frac{WB}{C}}\right)$$

$$\text{and} \quad m = \sqrt{\frac{W}{B}C}$$

putting $c_1/c_2 = C$.

Thus we get the optimum values of n and m.

If we wish to determine the optimum values of n and m for each stratum, we use (6.2.7) and (6.2.10).

Thus $\sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}(B_t + W_t/m_t) + \mu \sum_{t=1}^{k}(c_1 n_t + c_2 m_t n_t)$ should be minimum.

Minimisation gives

$$n_t = \lambda_t\sqrt{B_t/\mu c_1} \qquad m_t n_t = \lambda_t\sqrt{W_t/\mu c_2}$$

$$\text{i.e.} \quad m_t = \sqrt{\frac{W_t c_1}{B_t c_2}}$$

Putting for m_t and n_t in

$$\sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}(B_t + W_t/m_t) = V_0, \text{ we have}$$

$$\sum_{t=1}^{k} \frac{\lambda_t^2}{\lambda_t}\sqrt{\frac{\mu c_1}{B_t}}\left(B_t + W_t\sqrt{\frac{B_t c_2}{W_t c_1}}\right) = V_0,$$

$$\text{i.e.} \sqrt{\mu c_1}\sum_{t=1}^{k}\lambda_t\left(\sqrt{B_t} + \sqrt{W_t\frac{c_2}{c_1}}\right) = V_0,$$

$$\sqrt{\mu c_1} = \frac{V_0}{\sum_{t=1}^{k}\lambda_t\left(\sqrt{B_t} + \sqrt{\frac{W_t}{C}}\right)}$$

Therefore

$$n_t = \frac{\lambda_t \sqrt{B_t} \sum_{t=1}^{k} (\sqrt{B_t} + \sqrt{W_t/C})}{V_0}$$

6.2.12

$$\text{and} \quad m_t = \sqrt{\frac{W_t}{B_t} C} \quad .$$

It may be noted that m_t is used in place of m. What we do in practice is that we determine m_t for each stratum and then take the average as the value of m. We thus have determined the optimum values of n_t and m.

The calculation of sample size will be done on the basis of equation (6.2.11) but to illustrate the stratum-wise calculations, some cases will be worked out on the basis of equation (6.2.12).

6.3 SAMPLE SIZE

Note: n denotes the number of villages required to be sampled and m denotes the number of holdings required to be sampled in each selected village.

Table 6.1

Wheat

(i) Human Labour

m	5% S.E. n	mn	1% S.E. n	mn
1	67	67	1660	1660
2	44	88	1091	2182
3	36	108	901	2703
4	33	132	807	3228
5	30	150	750	3750
6	29	174	711	4266
28	28	196	685	4795
8	27	216	664	5312
9	26	234	649	5841
10	26	260	636	6360
11	25	275	623	6853
12	25	300	617	7404

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	2	43	86	1069	2138
2.5	3	41	123	1010	3030
3.0	3	39	117	968	2904
3.5	3	39	114	934	2802
4.0	3	37	111	907	2721

where $c = c_1/c_2$, c_1 being cost per village and c_2 the cost per holding.

Table 6.2

Wheat

Bullock Labour

m	5% S.E. n	mn	1% S.E. n	mn
1	94	94	2339	2339
2	54	108	1331	2662
3	40	120	995	2985
4	33	132	827	3308
5	29	145	726	3630
6	27	162	659	3954
7	25	175	611	4277
8	23	184	575	4600
9	22	198	547	4923
10	21	210	524	5240
11	21	231	506	5566
12	20	240	491	5892

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	4	36	144	894	3576
2.5	4	34	136	833	3332
3.0	5	32	160	789	3945
3.5	5	31	155	754	3770
4.0	5	29	145	726	3630

Table 6.3

__Wheat__

Total Input

m	5% S.E. n	mn	1% S.E. n	mn
1	87	87	2166	2166
2	51	102	1272	2544
3	39	117	973	2919
4	33	132	824	3296
5	30	150	734	3670
6	27	162	675	4050
7	26	182	632	4424
8	24	192	600	4800
9	23	207	575	5175
10	23	230	555	5550
11	22	242	539	5929
12	21	252	525	6300

__Optimum Sample Size (minimum cost):__

c	m	n	mn	n	mn
2.0	3	39	117	985	2874
2.5	4	36	144	896	3584
3.0	4	34	136	851	3404
3.5	4	33	132	815	3260
4.0	5	32	150	787	3935

Table 6.4

Wheat

Total Output

m	5% S7E. n	mn	1% S.E. n	mn
1	38	38	931	931
2	26	52	643	1286
3	22	66	547	1641
4	20	80	499	1996
5	19	95	470	2350
6	18	108	451	2806
7	18	126	438	3066
8	17	136	427	3416
9	17	153	419	3771
10	17	170	413	4130
11	17	187	408	4480
12	17	204	403	4836

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	2	27	54	676	1352
2.5	2	26	52	641	1282
3.0	3	25	75	617	1891
3.5	3	24	72	597	1791
4.0	3	24	72	581	1773

Table 6.5

Sugarcane (planted)

Human Labour

m	5% S.E. n	mn	1% S.E. n	mn
1	97	97	2404	2404
2	63	126	1555	3110
3	51	153	1272	3816
4	46	184	1130	4520
5	42	210	1045	5225
6	40	240	989	5934
7	38	266	948	6636
8	37	296	918	7344
9	36	324	894	8046
10	35	350	875	8750
11	35	285	860	9460
12	34	408	847	10164

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	3	60	180	1482	4446
2.5	3	56	168	1398	4194
3.0	3	54	162	1338	4140
3.5	3	52	156	1291	3873
4.0	4	51	204	1253	5012

Table 6.6

Sugarcane (planted)

Bullock Labour

m	5% S.E. n	mn	1% S.E. n	mn
1	179	179	4464	4464
2	97	194	2428	4856
3	70	210	1749	5247
4	57	228	1409	5636
5	49	245	1206	6030
6	43	258	1070	6420
7	39	273	973	6811
8	36	288	900	7200
9	34	306	844	7596
10	32	320	799	7990
11	31	341	762	8382
12	30	360	731	8772

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	5	52	260	1286	6430
2.5	6	48	288	1190	7140
3.0	6	45	270	1121	6726
3.5	6	43	258	1066	6393
4.0	7	41	287	1022	7154

Table 6.7

Sugarcane (planted)

Seed

m	5% S.E. n	mn	1% S.E. n	mn
1	40	40	984	984
2	28	56	683	1366
3	24	72	583	1749
4	22	88	533	2132
5	24	105	503	2515
6	20	120	483	2898
7	19	133	469	3283
8	19	152	458	3664
9	18	162	450	4050
10	18	180	443	4430
11	18	198	438	4818
12	18	216	433	5196

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	2	30	60	723	1446
2.5	2	28	56	687	1374
3.0	3	27	81	661	1983
3.5	3	27	78	639	1917
4.0	3	25	75	623	1869

Table 6.8

Sugarcane (planted)

Fertilizers and Manures

m	5% S.E. n	5% S.E. mn	1% S.E. n	1% S.E. mn
1	264	264	6588	6588
2	151	302	3762	7524
3	113	339	2820	8460
4	94	376	2348	9392
5	83	415	2066	10330
6	75	450	1877	11262
7	70	490	1743	12201
8	66	528	1642	13136
9	63	567	1563	14067
10	60	600	1501	15010
11	58	638	1449	15939
12	57	684	1406	16872

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	4	103	412	2566	10264
2.5	4	94	376	2390	9560
3.0	5	91	455	2264	11320
3.5	5	87	435	2164	10820
4.0	5	84	420	2085	10425

Table 6.9

Sugarcane (planted)

Total Input

	5% S.E.		1% S.E.	
m	n	mn	n	mn
1	55	55	1376	1376
2	36	72	902	1804
3	30	90	745	2235
4	27	108	666	2664
5	25	125	618	3090
6	24	144	587	3522
7	23	161	564	3948
8	22	176	547	4376
9	22	198	534	4806
10	21	210	526	5260
11	21	231	515	5665
12	21	252	508	6096

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	3	36	108	881	2643
2.5	3	34	102	832	2496
3.0	3	32	96	797	2391
3.5	3	31	93	770	2310
4.0	3	30	90	747	2241

Table 6.10

Sugarcane (planted)

Total Output

m	5% S.E. n	mn	1% S.E. n	mn
1	93	93	2318	2318
2	66	132	1632	3264
3	57	171	1403	4209
4	52	208	1289	5156
5	49	245	1220	6100
6	47	282	1175	7050
7	46	322	1142	7994
8	45	360	1104	8832
9	44	396	1099	9891
10	44	440	1083	10830
11	43	473	1071	11781
12	43	516	1060	12720

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	2	71	142	1754	3508
2.5	2	67	134	1667	3334
3.0	2	65	130	1604	3208
3.5	3	63	189	1555	4665
4.0	3	61	183	1516	4548

Table 6.11

Gram

Human Labour

m	5% S.E. n	mn	1% S.E. n	mn
1	155	155	3861	3861
2	89	178	2123	4246
3	62	186	1544	4632
4	51	204	1254	5016
5	44	220	1080	5400
6	39	234	964	5784
7	36	252	882	6174
8	33	264	820	6560
9	31	279	771	6939
10	30	300	733	7330
11	28	308	701	7711
12	27	324	675	8100

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	5	49	245	1205	6025
2.5	5	45	225	1117	5585
3.0	6	43	258	1054	6324
3.5	6	41	246	1003	6018
4.0	6	39	234	963	5778

Table 6.12

Gram

Bullock Labour

	5% S.E.		1% S.E.	
m	n	mn	n	mn
1	151	151	3777	3777
2	80	160	1995	3990
3	56	168	1401	4203
4	45	180	1104	4416
5	37	185	926	4630
6	33	198	807	4842
7	29	203	732	5061
8	27	216	659	5272
9	25	225	609	5481
10	23	230	570	5700
11	22	242	537	5907
12	21	252	510	6120

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	6	34	204	832	4992
2.5	7	31	217	765	5355
3.0	7	29	203	717	5019
3.5	8	28	224	680	5440
4.0	9	26	194	649	5841

Table 6.13

Gram

Total Input

m	5% S.E. n	mn	1% S.E. n	mn
1	77	77	1914	1914
2	42	84	1031	2062
3	30	90	736	2208
4	24	96	589	2356
5	20	100	501	2505
6	18	108	442	2652
7	16	112	400	2800
8	15	120	368	2944
9	14	126	344	3096
10	13	130	324	3240
11	13	143	308	3388
12	12	144	295	3540

Optimum Sample Size (minimum cost):

θ	m	n	mn	n	mn
2.0	5	21	105	509	2545
2.5	6	19	114	470	2820
3.0	6	18	108	442	2652
3.5	7	17	119	420	2940
4.0	7	16	112	402	2814

Table 6.14

Gram

Total Output

m	5% S.E. n	5% S.E. mn	1% S.E. n	1% S.E. mn
1	156	156	3887	3887
2	100	200	2494	4988
3	82	246	2030	6090
4	72	288	1797	7188
5	67	335	1658	8290
6	63	378	1565	9390
7	60	420	1499	10493
8	58	464	1449	11592
9	57	513	1410	12690
10	56	560	1379	13790
11	55	605	1354	14894
12	54	648	1333	15996

Optimum Sample Size (minimum cost):

c	m	n	mn	n	mn
2.0	3	94	282	2343	7029
2.5	3	89	267	2209	6627
3.0	3	85	255	2113	6339
3.5	3	82	246	2037	6111
4.0	4	79	316	1976	7904

6.4 <u>Stratum-wise estimation of the sample size</u>

Note: The percentage standard error is for the overall sample .

Table 6.15

<u>Wheat</u>

Human Labour

		5% S.E.						1% S.E.					
c	m	n_1	n_2	n_3	n_4	n	mn	n_1	n_2	n_3	n_4	n	mn
2.0	4	14	4	11	4	33	132	350	90	260	95	795	3180
2.5	4	14	4	10	4	32	128	330	84	245	89	748	2992
3.0	5	13	4	10	4	31	124	315	81	234	85	715	3575
3.5	5	13	4	10	4	31	124	312	80	231	84	707	3535
4.0	5	12	3	9	4	28	112	294	75	218	80	667	3335

Table 6.16

<u>Sugarcane (planted)</u>

Human Labour

2.0	3	12	5	21	20	58	174	284	118	525	501	1428	4284
2.5	3	11	5	20	19	55	165	268	112	496	473	1349	4047
3.0	3	11	5	19	18	53	159	256	107	474	453	1290	3870
3.5	4	10	5	19	18	52	156	247	103	458	437	1245	4980
4.0	4	10	4	18	17	49	147	240	100	444	424	1208	4832

Table 6.17

Sugarcane (planted)

Total Output

c	m	5% S.E. n_1	n_2	n_3	n_4	n	mn	1% S.E. n_1	n_2	n_3	n_4	n	mn
2.0	2	12	11	22	29	74	148	284	259	530	716	1789	3578
2.5	2	11	10	21	28	70	140	270	246	504	681	1701	3402
3.0	3	11	10	20	27	68	204	260	237	485	656	1638	4914
3.5	3	10	10	19	26	65	195	252	230	470	636	1588	4764
4.0	3	10	9	19	25	63	189	246	224	458	620	1548	4644

Table 6.18

Gram

Human Labour

2.0	5	14	5	16	5	40	200	331	117	389	107	944	4720
2.5	6	13	5	15	4	37	222	307	108	360	99	874	5244
3.0	7	12	4	14	4	34	238	289	102	339	94	824	5768
3.5	7	11	4	13	4	32	224	277	98	325	90	790	5530
4.0	8	11	4	13	4	32	256	264	93	310	86	753	6024

7. EFFICIENCY OF STRATIFICATION

7.1 The four-th part of the study concerns the gain in precision of the estimates and usefulness or otherwise of the stratification at the ultimate sampling stage as adopted in the survey.

In each selected village all the holdings were first arranged in the order of the total area of holding. The holdings were then divided into five equal groups, the Ist one-fifth of the ordered holdings forming the first group, the second one-fifth the second one and so on. From each group, two holdings were selected at random and thus a sample of ten holdings was selected from the entire village. To find out whether such stratification leads to any gain in precision of the estimate of the population mean, we evaluate the gain first theoretically and then estimate the same from the sample selected. The gain in precision is the variance of the estimate of the population mean without taking into consideration the stratification of the holdings minus the variance when the stratification is taken into account. The procedure is as follows:

Let z_{tiuj} represent the j-th holding in the u-th group of the i-th village belonging to the t-th stratum.

Let L_{ti} be the number of groups in (t,i)th village,

M_{tiu} be the number of holdings in the u-th group of the (t,i)th village,

m_{tiu} be the number of holdings selected from the u-th group of the (t,i)th village.

The rest of the symbols used in this section bear the same meaning as in the previous section.

Considering the sample of holdings selected from a village as a simple random sample, we have the variance of the sample mean, $\bar{z}$, given by

7.1.1 $$V(\bar{z})_{us} = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}\left[\sigma_{tb}^2 + \sum_{i=1}^{N_t} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) S_{ti}^2 \right]$$

7.1.2 where $$\begin{cases} \sigma_{tb}^2 = \frac{1}{N_t - 1} \sum_{i=1}^{N_t} (z_{ti.} - \bar{z}_{t..})^2 \\ S_{ti}^2 = \frac{1}{M_{ti} - 1} \sum_{j=1}^{M_{ti}} (z_{tij} - \bar{z}_{ti.})^2 \end{cases}$$

whereas the same expression when the sample is considered a stratified sample, becomes

7.1.3 $$V(\bar{z})_s = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}\left[\sigma_{tb}^2 + \sum_{i=1}^{N_t} P_{ti} \sum_{u=1}^{L_{ti}} \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}} \right) p_{tiu}^2 S_{tiu}^2 \right]$$
$$= \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}\left[\sigma_{tb}^2 + \sum_{i=1}^{N_t} P_{ti} \sum_{u=1}^{L_{ti}} \left(\frac{p_{tiu}}{m_{tiu}} - \frac{1}{M_{ti}} \right) p_{tiu} S_{tiu}^2 \right]$$

where

7.1.4 $$p_{tiu} = M_{tiu}/M_{ti}$$
$$S_{tiu}^2 = \frac{1}{M_{tiu}-1} \sum_{j=1}^{M_{tiu}} \left(z_{tiuj} - \bar{z}_{tiu.} \right)^2$$

To estimate the gain in precision we have to break up S_{ti}^2 into two components viz. between groups mean square and within groups mean square.

Now $$\sum_{u=1}^{L_{ti}} \sum_{j=1}^{M_{tiu}} (z_{tiuj} - \bar{z}_{ti..})^2 = \sum_{u=1}^{L_{ti}} \sum_{j=1}^{M_{tiu}} (z_{tiuj} - \bar{z}_{tiu.} + \bar{z}_{tiu.} - \bar{z}_{ti..})^2$$

i.e. $$(M_{ti} - 1)\, S^2_{ti} = \sum_{u=1}^{L_{ti}} (M_{tiu} - 1)\, S^2_{tiu} + \sum_{u=1}^{L_{ti}} M_{tiu} (\bar{z}_{tiu.} - \bar{z}_{ti..})^2$$

where $\bar{z}_{tiu.}$ is the population mean of the u-th group of the i-th village in the t-th stratum and similarly for $\bar{z}_{ti..}$ etc.

Assuming $\frac{M_{ti}-1}{M_{ti}} \simeq 1$ and $\frac{M_{tiu}-1}{M_{ti}-1} \simeq \frac{M_{tiu}}{M_{ti}}$,

we have

7.1.5 $$S_{ti}^2 = \sum_{u=1}^{L_{ti}} p_{tiu} S_{tiu}^2 + \sum_{u=1}^{L_{ti}} p_{tiu} \left(\bar{z}_{tiu.} - \bar{z}_{ti..} \right)^2$$

As the groups are all equal, we have $m_{tiu} = p_{tiu} m_{ti}$ i.e. we have proportionate sampling.

Thus (7.1.3) becomes

$$V(\bar{z})_s = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}\left[\sigma_{tb}^2 + \sum_{i=1}^{N_t} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) \sum_{u=1}^{L_{ti}} p_{tiu} S_{tiu}^2 \right]$$

7.1.6 $$= \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t}\left[\sigma_{tb}^2 + \sum_{i=1}^{N_t} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) \left\{ S_{ti}^2 - \sum_{u=1}^{L_{ti}} p_{tiu} (\bar{z}_{tiu.} - \bar{z}_{ti..})^2 \right\} \right]$$

with the help of (7.1.5).

Subtracting (7.1.6) from (7.1.1), we have the gain given by

7.1.7 $$V(\bar{z})_{us} - V(\bar{z})_s = \sum_{t=1}^{k} \frac{\lambda_t^2}{n_t} \sum_{i=1}^{N_t} P_{ti} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}} \right) \sum_{u=1}^{L_{ti}} p_{tiu} \left(\bar{z}_{tiu.} - \bar{z}_{ti..} \right)^2$$

The estimate of (7.1.7) will give us the estimate of the gain in precision and for this we have to estimate $\sum_{u=1}^{L_{ti}} p_{tiu} (\bar{z}_{tiu.} - \bar{z}_{ti..})^2$

7.2 To estimate $\sum_{u=1}^{L_{ti}} p_{tiu} (\bar{z}_{tiu.} - \bar{z}_{ti..})^2$, we proceed as follows:

Let

7.2.1 $$\bar{z}_{tiu(m_{tiu})} = \bar{z}_{tiu.} + \epsilon_u$$

(Note that $Z_{tiu.}$ is the population value whereas $Z_{tiu(m_{tiu})}$ is the sample estimate)

7.2.2 where $E(\epsilon_u) = 0$ and

$$E(\epsilon_u^2) = \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) S^2_{tiu}.$$

Squaring both sides of (7.2.1), we have, taking expectations,

$$E\left[\bar{z}^2_{tiu(m_{tiu})}\right] = \bar{Z}^2_{tiu.} + \left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) S^2_{tiu}$$

7.2.3 $$\therefore E\left[\sum_{u=1}^{L_{ti}} p_{tiu}\bar{z}^2_{tiu(m_{tiu})}\right] = \sum_{u=1}^{L_{ti}} p_{tiu}\bar{Z}^2_{tiu.} + \sum_{u=1}^{L_{ti}} p_{tiu}\left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) S^2_{tiu}.$$

Transposing, we have

7.2.4 $$Est.\sum_{u=1}^{L_{ti}} p_{tiu}\bar{Z}^2_{tiu.} = \sum_{u=1}^{L_{ti}} p_{tiu}\bar{z}^2_{tiu(m_{tiu})} - \sum_{u} p_{tiu}\left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) s^2_{tiu}$$

$$\because Est\ S^2_{tiu} = s^2_{tiu}$$

Also we have, from (7.2.1)

$$\sum_{u=1}^{L_{ti}} p_{tiu}\bar{z}_{tiu(m_{tiu})} = \sum_{u=1}^{L_{ti}} p_{tiu}\bar{Z}_{tiu.} + \sum_{u=1}^{L_{ti}} p_{tiu}\epsilon_u$$

7.2.5 $$\therefore \bar{z}_{ti(m_{ti})} = \bar{Z}_{ti..} + \sum_{u=1}^{L_{ti}} p_{tiu}\epsilon_u.$$

Squaring both sides, we have $\bar{z}^2_{ti(m_{ti})} = \bar{Z}^2_{ti..} + \sum_{u=1}^{L_{ti}} p^2_{tiu}\epsilon^2_u + \sum_{u\neq u'} p_{tiu}p_{tiu'}\epsilon_u\epsilon_{u'} + 2\bar{Z}_{ti..}\sum_{u=1}^{L_{ti}} p_{tiu}\epsilon_u$

Taking expectations, we have, putting $E(\epsilon_u\epsilon_{u'}) = 0$, as u and u' are independent,

7.2.6 $$E\left[\bar{z}^2_{ti(m_{ti})}\right] = \bar{Z}^2_{ti..} + \sum_{u=1}^{L_{ti}} p^2_{tiu}\left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) S^2_{tiu}.$$

7.2.7 $$\therefore Est\ \bar{Z}^2_{ti..} = \bar{z}^2_{ti(m_{ti})} - \sum_{u=1}^{L_{ti}} p^2_{tiu}\left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) s^2_{tiu}.$$

Subtracting (7.2.7) from (7.2.4), we have

7.2.8 $$Est\sum_{u=1}^{L_{ti}} p_{tiu}(\bar{Z}_{tiu.} - \bar{Z}_{ti..})^2 = \sum_{u=1}^{L_{ti}} p_{tiu}\left\{\bar{z}_{tiu(m_{tiu})} - \bar{z}_{ti(m_{ti})}\right\}^2 - \sum_{u=1}^{L_{ti}} p_{tiu}(1-p_{tiu})\left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) s^2_{tiu}.$$

Substituting in (7.1.7), we have the estimate of the gain as

$$\sum_{t=1}^{K}\frac{\lambda_t}{n_t}\sum_{i}^{n_t}\left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}}\right)\left[\sum_{u=1}^{L_{ti}} p_{tiu}\left\{\bar{z}_{tiu(m_{tiu})} - \bar{z}_{ti(m_{ti})}\right\}^2 - \sum_{u=1}^{L_{ti}} p_{tiu}(1-p_{tiu})\left(\frac{1}{m_{tiu}} - \frac{1}{M_{tiu}}\right) s^2_{tiu}\right]$$

which (when M_{ti} and M_{tiu} are large) may be written as

7.2.9 $$\sum_{t=1}^{K}\frac{\lambda_t^2}{n_t}\sum_{i}^{n_t}\frac{1}{m_{ti}}\left[\sum_{u=1}^{L_{ti}} p_{tiu}\left\{\bar{z}_{tiu(m_{tiu})} - \bar{z}_{ti(m_{ti})}\right\}^2 - \sum_{u=1}^{L_{ti}} p_{tiu}(1-p_{tiu})\frac{s^2_{tiu}}{m_{tiu}}\right]$$

7.3 Having obtained the expression for the gain in precision of the final estimate due to stratification within a village, we proceed to actually calculate the same for some of the major factors e.g. Human Labour/acre, Bullock Labour/acre, Total Input/acre and Total Output/acre for Wheat, Sugarcane (planted) and Gram.

The figures for the above factors are given as follows :

	Crop	Factors (per acre)	Percentage gain
(i)	Wheat	Human Labour	1.40
(ii)	"	Bullock Labour	6.71
(iii)	"	Total Input	5.40
(iv)	"	Total Output	0.81
(v)	Sugarcane (planted)	Human Labour	2.08
(vi)	"	Bullock Labour	7.05
(vii)	"	Total Input	1.16
(viii)	"	Total Output	2.13
(ix)	Gram	Human Labour	0.00
(x)	"	Bullock Labour	0.00
(xi)	"	Total Input	0.00
(xii)	"	Total Output	27.88

7.4 From the above figures we observe that, on the average, there is practically no gain due to stratification of holdings within a village. We may thus conclude that stratification as adopted in the present survey does not yield any gain in precision. Further, it also does not help in the estimation of values for specific size-classes of holdings, since the range of strata sizes is not fixed but varies from village to village and adjustment for these variable strata complicates the estimation. The possibility of stratification of holdings within villages according to predetermined size ranges, e.g. < 2.5, < 5.0 acre etc. needs to be explored in order to facilitate estimation according to specific strata sizes.

S U M M A R Y

8.1 Although the studies on the various aspects of farm management started in the Western countries in the beginning of the present century and have since been getting considerable importance, in India no general efforts seem to have been made in this direction. Whatever data were collected, contained information on some specific items only. No broad-based effort was made till very recently when the scheme 'Studies in Economics of Farm Management' was started, with the objective of collecting primary data on various cost and production aspects of the farm management and for comparison of alternative methods of collecting the data.

8.2 The present study, (described in the previous sections,) deals with four aspects of this :

1. In an effort to find out the functional relationship existing between the factors of input on the one hand and output on the other, various production functions have been discussed. Of these, the Cobb-Douglas production function has been fitted to all the crops although, sometimes, the quadratic function has also been tried. We do not succeed in arriving at a close relationship for the reasons that

(i) the first year data (used here) may not be of the required standard and quality,

(ii) the factors considered in the various functions may not be the ones actually contributing towards the production and that there are some 'extraneous' factors e.g. climate, soil-fertility etc. which need be studied.

2. In the case of comparison between the Cost Accounting

and Survey methods of collecting the primary data, the comparison was made in two ways. The first type of comparison was made on the basis of independent villages and the second one on the basis of common villages. The two types of comparison do not yield identical results, there being more results significant in the latter type which is to be expected; there is no contradiction in the two types of comparison i.e. they give similar results. The methods were found to differ in almost all the factors but the difference is quite marked in the case of Bullock Labour/acre and Total Input/acre, the excess being in favour of the Survey method.

3. In the case of estimation of sample size for the various strata individually as well jointly, sample sizes have been calculated for different standard errors viz. 5% and 1%.

4. In the case of estimation of gain in precision and usefulness of the stratification of holdings in a village as adopted in the survey, it has been found that there is practically no gain in the mode of stratification within the primary sampling units (as adopted).

The possibility of the usefulness of stratification according to fixed size ranges needs to be explored.

Acknowledgements

I am deeply indebted to Dr. V.G. Panse, Statistical Adviser, Indian Council of Agricultural Research, under whose guidance this dissertation has been prepared. But for his inspiration and constant encouragement, it would not have been possible to present the dissertation in this form.

I am also grateful to Dr. B.V. Sukhatme and Shri Daroga Singh who have been giving invaluable help and many useful suggestions from time to time.

BIBLIOGRAPHY

1. Bhattacharjee, J.P. and Sharma, J.S. — An account of investigation in the field of farm management studies in India 1957.

2. Bhattacharjee, J.P. and Sharma, J.S. — Report on the F.A.O. farm management development centre for the Asia and the Far East, Tokyo. 1956.

3. Bradford, Lawrence A. and Johnson, Glenn L. — Farm Management Analysis. John Wiley & Sons, 1953.

4. Daniel, E.F. — The role of farm management in agricultural extension service. Indian Jour. of Agronomy, vol.II, No.1, 1951.

5. Das, M.N. — Analysis of covariance in two-way classification with disproportionate cell frequencies. Jour. of the Indian Soc. of Agrl. Stat. vol.V, No.2, 1953.

6. Heady, Earl O. — Economics of Agricultural Production and Resource Use. Prentice - Hall Inc. New York, 1952.

7. Heady, Earl O. — Production function for a random sample of farms. Jour. of Farm Econ. vol.XXVIII, No.4, Nov. 1946.

8. Iback, Donald B. — Production functions in management research Jour. of Farm Econ. vol. XXXV, No.5, Dec. 1953.

9. Johnson, Paul R. — Alternative functions for analysing fertiliser-yield relationship. Jour. of Farm Econ. vol. XXXV, No.4, Nov. 1953.

10. Juichi Sato — Programmes prevailing in Japan for extension services in the fields of farm management techniques and research. 1957.

11. Kendall, M.G. — The Advanced Theory of Statistics. Charles Griffin & Co., London, 1951.

12. Lewis, A.B. — Farm management programmes and policies in the Far East. 1957.

13. Panse, V.G. — Estimation of the cost of production of crops. Indian Central Cotton Committee, Bombay, 1954.

14. Panse, V.G. — Some observations on the problems of estimating farm costs. 1957.

15. Rao, C.R. Advanced Statistical Methods in Biometric Research. John Wiley & Sons, New York, 1952

16. Rao, K.S. Some techniques of processing farm management data. 1957.

17. Rodman, John C. Problems and possible solutions in determining input-output relationships in agricultural production. Jour. of Farm Econ. vol. XXXVI, No.5, Dec. 1954.

18. Sen, S.R. Importance of farm management studies in stepping up agricultural production. 1957.

19. Snedecor, George W. Statistical Methods. Collegiate Press Inc., Iowa, 1946.

20. Sukhatme, P.V. Sampling Theory of Surveys with Applications. Indian Soc. of Agrl. Stat., New Delhi, India;

21. Titner, G. A note on the derivation of production function from farm records. Econometrica, 1944.

22. Zvi Griliches Specification bias in estimates of production functions. Jour. of Farm Econ., vol.XXXIX, No.1, Feb. 1957.