Proceedings of the Symposium on Utilisation of Available Data in Agricultural Research

Chairman : Dr. O.P. Kathuria
Director, IASRI, New Delhi.

Convenors : Dr. A.K. Srivastava and
Principal Scientist &
Head, Div of SSM
IASRI, New Delhi

Prof. P.V. Prabhakaran
Head, Dept. of Statistics,
College of Horticulture,
KAU, Trichur.

This symposium was organised during 48th Annual Conference of the Indian Society of Agricultural Statistics on 16 December, 1994 at the College of veterinary and Animal Sciences, Kerala Agricultural University, Mannuthy, Thrissur. The Chairman thanked the Society and organisers for inviting him to preside over the session. He emphasised the importance of the topic and its role in the present context of rapid growth of agricultural research. The complexities of problem in agricultural research, in view of growing emphasis on multidisciplinary approach, have increased manifold. The role of data collection and its analysis in every aspect of agricultural research needs no emphasis. At the same time the cost aspect of data collection need not be lost sight of. In every area of agricultural research data is generally collected and utilised keeping in view the specific objectives of the study. Many a times, the data collected has got much more potential than what is harnessed from it. It is also not uncommon to find that data is not easily accessible to many potential users. All these aspects provide bottlenecks in the optimum utilisation of available data. The Chairman emphasised that it is in this context that the present topic becomes most relevant. After these introductory remarks, the Chairman invited the speakers to present their papers. In all, eight papers were presented covering analysis of data in various areas of agricultural research such as forestry, marine fisheries, tuber crops, plantation crops, dairy cattle, fisheries technology, and agricultural surveys.

Based on the presentation of papers, subsequent discussions and observations of the Chairman, the following recommendations were made :

- Since most of the data analysis is computerised, the data collected through experimentation and surveys, should be properly documented which may ultimately lead to generation of data bases.

- Analytical studies of data should be encouraged. The Society may contribute significantly in this respect through taking up need based analytical studies in the form of research projects.
- Multidisciplinary studies utilising data from various sources should be taken up. For this purpose linkages of research workers in statistics in various Institutes should be further strengthened.

The extended summaries of the papers are as follows :

1. Trend Analysis in Plantation Crops

M.V. George
Central Plantation Crops Research Institute, Kasaragod

India is the largest producer of arecanut, cashewnut, tea, ginger and turmeric and second largest producer of coconut, cardamom and pepper in the world. Since independence an increase in trend in the area and production of these crops were seen. However the trend in area or production are neither uniform among the crops nor over the periods, as the same were very much influenced by the climatic and seasonal changes, market prices, occurrence of drought and cyclones, out break of disease and pest incidence. This paper dwells at length the trend in area, production and productivity of all these crops for the past four decades, utilising the available data.

Coconut

There was an overall increase of 160 percent in area and 218 percent in production during 1992-93 compared to 1950-51 figures. The growth rates in area slided down to nearly zero during the seventies and further improved during eighties and in particular from 1986 onwards. The increase in production for the last ten years alone was about 96%. The compound growth rate for 1984-90 and 1990-93 were 9.85 and 8.29 respectively. The average productivity of 5759 nuts per ha during 1950-51 rose to7032 during 1992-93. The compound growth rate which was negative during sixties and seventies became positive during late eighties and nineties.

Arecanut

Area and production trend of arecanut in India increased during the past 38 years. The area increased by 110% and production by 205%. The compound growth rate for area as well as production were generally positive and significant. The productivity of arecanut remained almost steady from 1957 to

1974 with small fluctuations, then increased steadily to above 1000 kg/ha level and then reached the peak level of 1250 kg/ha during the year 1989-90.

Cashew

The area under cashew showed an increasing trend. There was an overall increase of 414 percent in area and 341 percent in production during 1993-94 when compared to 1955-56. The compound growth rates were all positive, high and significant. The productivity was showing a declining trend till the end of seventies. The highest productivity 768.2 kg/ha observed during 1957-58 had declined to 465 kg during 1975-76 and further to the lowest level of 314.27 kg/ha during 1979-80, about 60% reduction as compared to the level of productivity during 1957-58. The productivity increased from 452.35 kg/ha during 1985-86 and to 615.98 kg/ha during 1993-94.

Cocoa

The area under cocoa was showing a negative trend from the year 1983-84. The area under cocoa during the year 1983-84 was 22230 ha which came down to 13960 ha during 1991-92 whereas the production was fluctuating between 56100 and 77100 kg. However the productivity was almost increasing every year. The compound growth rate for area was negative and significant, but that of production was not significant. However, CGR of productivity was positive and significant.

Pepper

The area under pepper in India which was only 79000 ha during 1949-50 increased to 177360 ha during 1992-93. The compound growth rate of area was positive, high and significant. The production of pepper in India increased from 21000 tonnes during 1949-50 to 55190 tonnes during 1989-90. During the past 10 years i.e. from 1983-84 to 1992-93 the increase in production was around 138 percent and the compound growth rate was highest during the five year period 1984-89. As regards the productivity, it was stagnant within the range of 250-300 kg/ha with marginal fluctuations during the past 44 years. The lowest productivity was observed during the year 1984-85 mainly due to spread of quick wilt and loss of pepper wines in extensive areas consequent to the heavy and continuous rains in Kerala. However the productivity picked up from 1987 onwards.

Cardamom

The area under cardamom was showing an upward trend from 1955-56 to 1993-94 with occasional ups and downs. However the expansion of area was only 66 percent over the 38 years. The production was characterised by an overall increasing trend between 1955-56 and 1993-94, although the production was stagnant for few years. The trend in productivity was found highly fluctuating, ranging from 24 kg to 84 kg/ha without showing any regular trend.

Coffee

The area under coffee during 1950-51 was only 0.091 million ha and increased to 0.279 million ha during 1991-92. The production of 0.034 million tonnes during 1955-56 increased to 0.18 million tonnes during the year 1991-92. The compound growth rates for area as well as for production were all positive and significant. The productivity was highly fluctuating.

Tea

The area under and production of tea in India was showing an increasing trend. The area during 1956 was only 0.316 million ha which increased to 0.412 million ha during 1986 and then remained stationary for the last few years. The production of 0.285 million tonnes during the year 1956 increased to 0.715 million tonnes during 1990. The productivity was also showing an upward trend with ups and downs. The productivity of tea during the year 1956 was 902 kg/ha and increased to 1787 during 1990. The compound growth rate of area worked out for the five years period was positive and significant upto the period 1980-85 and there after it was negative. But the CGR of production was positive throughout and mostly significant. The CGR for productivity was also positive.

Rubber

The area under and production of rubber was in the increasing trend. The area under rubber during the year 1950-51 was only 0.059 million ha and rose to 0.5084 million ha during 1993-94. The production of rubber which was only 0.014 million tonnes during 1950-51 rose to 0.435 million tonnes during 1993-94, i.e. about 31 times increase over a period of 44 years. The compound growth rate of area and production for the five years period were all positive, high and significant. The productivity of rubber during 1950-51 which was 237 kg/ha increased to 856 kg/ha during 1993-94.

Ginger

The area under ginger was found to be fluctuating around 20,000 ha between 1949-50 to 1973-74. However, since 1974 an increasing trend was found. The area under ginger during 1972-73 was only 22870 ha, had increased to 58080 ha during the year 1992-93, about 154% increase over a period of 20 years. The production was about 24000 tonnes during 1949-50 and rose to the peak level of 189440 tonnes during 1992-93. The increase in production from 1983-84 to 1992-93 was of the order 84 percent. The productivity of ginger was found to be highly fluctuating around 1000 kg/ha or below from the year 1950-51 to 1969-70 thereafter it was showing an upward trend. The productivity which was 952 kg/ha during 1967-68 rose to 3262 kg/ha during 1992-93. As compared to fifties and sixties the productivity of ginger trebled during the nineties.

Turmeric

The area under turmeric remained around 50000 ha with little ups and downs from 1949-50 to 1961-62 and thereafter it was found increasing. The area rose from 68400 ha during 1972-73 to 128600 during the year 1992-93. The production of turmeric was mostly in the range of 130 to 150000 tonnes without much improvement from 1949-50 to 1978-79 and thereafter it had recorded substantial increase. It attained the peak level of 397400 tonnes during the year 1992-93, registering an increase of about 105 percent during the last 10 years. The productivity of turmeric did not improve much when compared to 1949-50.

Future Prospects

The overall trend in area and production of all these crops are positive and significant. From the trend analysis, it is estimated that the production of coconut by 2000 AD may reach 20000 million nuts from an estimated area of 2.1 million ha. The production of arecanut is estimated to exceed 3 lakh tonnes from an estimated area of 2.5 lakh ha. The cashew production is expected to reach 7 lakh tonnes by 2000 AD from an estimated area of 7 lakh ha. As per the trend the area under cocoa is not likely to increase. For pepper and cardamom the positive trend in area as well as production is likely to continue. By 2000 AD the pepper production is expected to reach one lakh tonnes from an estimated area of about 3 lakh ha. For cardamom though there is not much scope for area expansion there is scope for increasing the productivity of 81 kg/ha to 90 kg/ha and the estimated production of cardamom will be around 7000 tonnes. For coffee the production is likely to exceed 2.25 lakh tonnes from an estimated area of 3.18 lakh ha. For tea the scope for the expansion of area is very limited. The production may reach 7.5 lakh tonnes by increasing the productivity. The positive and significant rate of increase in area, production and productivity of rubber is likely to continue. The estimated production of rubber will be around 6 lakh tonnes from an area of 6 lakh ha. For ginger and turmeric also considerable expansion in area and increase in production is expected.

2. Utilization of Available Data in Experiments with Plantation Crops

Jacob Mathew
Central Plantation Crops Research Institute, Kayangulam - 690 533

Special features of perennial plantation crops make experimentation difficult in such crops. Also because of the limitations of time, land etc. there are restrictions on starting new experiments, as and when required. Therefore, researchers are forced to critically look into old experimental data, to draw information regarding points, not originally envisaged, while starting the experiment. Some such instances, where old experimental data have been made use of to obtain information about optimum plot size, competition and border effect, method of compilation of annual yield, crop losses etc. are discussed below.

Uniformity trial data and optimum plot size.

Uniformity trials are intended to get an idea about the nature and extent of fertility variation in land. From the data collected from such trials, fertility contour maps showing lines passing through areas of equal fertility are generally prepared. In most cases, this variation in fertility, optimum size and shape of plots and blocks are determined. In plantation crops like coconut and oil palm, where the palm to palm distance is 7 to 9 m., in order to have a uniformity trial, at least 2 ha of land will be required. Usually it will be difficult to get this much of area under uniform management, and with individual tree yield records. Koch and Rogney (1951) and Ray et al. (1973) have suggested techniques to obtain uniformity trial data from experimental data, by eliminating the treatment effects, from the yield of individual palms. The resultant data are considered as uniformity trial data, and is used to estimate optimum size and shape of plots. Nambiar (1986, 1989) has used fertilizer trial data to estimate the optimum size of plots in the case of DXT and W.C. Tall palms. This technique will hold good only when the number of palms per plot is constant, and the yield figures are available for individual palms.

Border/competition effect

Owing to the border effect, the yield or other characters of the plants near the borders differ from those at the centre of the plot. To overcome these problems, non-experimental guard rows are usually suggested. It can be easily seen that when gross plot is smaller in size, a great proportion of the experimental area would be discarded as non-experimental margins. Gomez and Gomez (1984) have mentioned that competition effect can arise due to non-planted borders, varietal competition, fertilizer competition and missing plots, and have suggested experiments to measure competition effects. They have also suggested techniques to estimate this competition effects from the data available from other experiments. Data from an irrigation-cum-fertilizer-cum-variety experiment was made use of to study the necessity or otherwise of external guard rows and guard rows between unirrigated and irrigated plots in coconut field experiments (Jacob Mathew et al. unpublished). Similarly, data from a cocoa+areca mixed cropping experiment was used by Jose et al. (1993) to study intra and inter component competition.

Crop loss assessment

Crop loss assessment requires the collection of large amount of data in order to understand the interaction of factors that may be affecting the yield. These data need to be collected through surveys and field experiments. Field experiments tend to be more limited in scope, providing data on few specific factors affecting particular crop, and their relationship to that crop yield. Recently data collected from a fertilizer experiment was made use of to study the impact of root (wilt) disease on the yield of young coconut palms (Jacob Mathew et al. 1993). Data from some of the field experiments in the disciplines of crop improvement and crop production have also been made use to study the extent of bienniality in different varieties of coconut (Jacob Mathew & Jose 1990) method of compilation of annual yield (Jacob Mathew et al. 1990) etc. Thus it is clear that in the case of plantation crops, the necessity is more and the possibilities are also plenty for the use of data available from different experiments, for purposes other than what was originally envisaged for, while starting the experiment.

3. Data Analysis in Forestry Sector

K. Jayaraman

Kerala Forest Research Institute, Peechi, Trichur

Certain instances in forestry where conventional sampling, design and analytical techniques are not sufficient, are briefly referred to with indication on possible solutions that have been offered. The problems covered pertain to the estimation of animal abundance in forests through line transact sampling, projection of stand structural changes in forests through simulation, analysis of repeated measurements through univariate mixed model/multivariate analysis and also the construction of a multivariate selection index using regression with indicator response variable.

Estimation of animal abundance in forests

Most of the conventional sampling schemes deal with populations of stationary objects for which a sampling frame can be prepared. In the case of wildlife, the sampling units are not stationary and the very objective in most cases will be to estimate the total number of units in the population. The procedures get even more difficult with individuals which occur in low density, which are often difficult to see (nocturnal, shy or flee or hide), which can fluctuate widely in numbers and distribution from season to season and year to year, and can sometimes fight back. Certain specialized methods have been evolved to deal with the situation like capture-recapture technique, line transact sampling, bounded count method and methods based on indirect evidences like hoof marks, pugmarks, pellet or dung counts. Of the several methods, line transact sampling has been found to have several advantages by way of ease of execution in the field and by being a direct sighting method. Of late, the theoretical basis of this method has been thoroughly investigated and several advancements have been made in the estimation procedures.

Projection of structural changes in forests

Forests are complex dynamic entities undergoing continuous structural modifications by way of natural regeneration, survivor growth and mortality. In managed forests, there is the additional human intervention in the form of periodical harvests. Natural forests consist of trees of several species of different age/size classes competing with each other for their survival. Forests, natural or man-made have long gestation period and studies on the effect of alternative management strategies on such systems are not amenable to conventional experimentation based on field observations over long periods. Instead, modelling and simulation techniques are resorted to, for answering questions related to the optimum management schedules. A typical mode based on transition matrices is given by Buongiorno and Michie (1980).

Analysis of repeated measures

The case of repeated measurements is a very common situation in forestry. The term 'repeated' is used to mean measurements which are made of the same characteristic on the observational unit but on more than one occasion. Typical examples are periodical measurements on diameter, height and crown size of trees in a silvicultural trial. Repeated measurements may be spatial rather than temporal. For instance, consider measurements on wood characteristics of several stems at bottom, middle and top portion of each stem, each set of stems coming from a different species. Another example is concerned with soil properties observed from multiple core samples at 0-20 cm, 20-40 cm, 40-60 cm depth from different types of vegetation. Different ways of analysing repeated measures have been proposed in the past (Crowder and Hand, 1990). These methods vary in their efficiency and appropriateness depending upon the nature of data. In routine analysis, repeated measurements over time or space within an individual are simply taken as levels of a factor and analysis is done ignoring the possible within subject correlations. This is not a satisfactory procedure. Certain more refined procedures are available.

Univariate mixed model analysis of variance

[See Rouanet & Lepine (1990); Mauchly (1940); Box(1949)]

Multivariate analysis of variance

[See Crowden & Hand (1990); Morrison (1970)

A multivariate selection index for vegetative propagation

Multivariate selection and ranking procedures are usually based on Mahalanobis Distance function or linear combination of the elements of mean vector (Krishnaiah, 1966). But there could be instances where functions other than the above are to be used in selection procedures. These are discussed in the paper.

4. Utilisation of Survey Data in Fisheries Technology

H. Krishna Iyer

Central Institute of Fisheries Technology Cochin - 682029

Data are essential for planning, management and development in all fields, especially in Fisheries. The Central Institute of Fisheries Technology has made many studies involving survey and collection of data in connection with fishing and fish processing industry. Some of the surveys undertaken in the past are indicated. The analysis of data were carried out keeping in view the objectives. However, these data can be examined again to have some deeper analysis or to extract additional information.

Survey on the idle capacity of fish processing Plants in India

CIFT has made in 1982 an all India survey to estimate the idle capacity of plants existing in the different maritime states and the results are already published. Among states in the west coast of India, idle capacity was less in Kerala (67.4%) while in the east coast, Tamil Nadu registered less idle capacity (57.1%). It is time for estimating the idle capacity of the plants again, as aquaculture is getting momentum in the country. While estimating the idle capacity of plants, it is better if it is estimated season-wise.

In order to evaluate the economics of operation of trawlers, especially 9.75 m and 10.97 m, data on the cost of construction of the hull, engine, gear and accessories, crew wages and commissions, fuel, including engine oil and lubricating oil, the number of trips performed, the quantity of prawn and fish caught, total receipts, insurance charges, repairs and maintenance of the boat including spare parts were collected from the selected boats. Profit in relation to number of days of operation for 9.75 m and 10.97 m trawlers were studied. A study on the low energy fishing techniques comprising of gill netters and long liners was undertaken during 1990 in order to assess their economic performance. Data on capital cost of the boat, engine and gear, the variable cost of operation, total number of fishing trips made during the year and the catch composition were collected. The performance indicators such as marginal cost, average cost, the break even point and the percentage profit of the selected boats were examined.

An all India survey was undertaken during the year 1992 to study the supply and demand positions of fish nets in the country. The objectives were to assess the status of net making industry in India both in the artisanal and organised sectors covering the marine and inland states including the union territories, to get estimates of supply and demand of fish nets and to estimate the additional units of the net making machines needed to the installed in the country to meet the demand. In addition to this many experimental data in connection with harvest and post-harvest technology of fish were collected and analysed for technological research programmes and management of fisheries.

5. Marine Fishery Data Base and Utilisation

K. Alagaraja
Central Marine Fisheries Research Institute, Cochin - 682014

Assessing the resource potentials being the major thrust area of research of the Institute (CMFRI), emphasis is laid on the quality of data collected on marine fish landings. For this purpose the well tested stratified multistage random sampling design developed over the past few decades is used. The startification is over space and time. Further stratification is introduced in the type of gears operated for fishing.

Seasonal variations are encountered in the marine fish landings. To understand their fluctuations and to study the impact of environment on exploited fish stocks season-wise estimates are required. For this purpose a month is considered as the time stratum for estimation purposes. Quarter-wise estimates and annual estimate are built up from monthly estimates. Gearwise catches are particularly important to identify the gears most suited to catch certain species of fishes. Information on gear wise landings is needed for fish stock assessment studies. In order to evaluate the availability of species in our waters quality wise and quantity wise information is needed. Season wise, size wise, quantity wise availability of species helps in formulating suitable management and conservation policies for fishing valuable resources.

Data on the biology of exploited fish are required covering the information on growth, mortality, size at first maturity, size at first capture, season wise and size wise availability and sex wise abundance and the areas of their availability. Synchronising with the collection of other details of exploited stocks data on biology are also collected regularly and systematically.

Information on the different aspect of environment such as water currents, temperature, salinity, productivity and pollution is collected to find out their impact on the exploited fish stocks. Remote sensing data on sea surface temperature (SST) and plankton bloom obtained through satellite imageries are obtained for this purpose.

Location testing for site selection for pisciculture operation is another area where information is collected on chemical and physical nature of soil and water suited for fish culture. Data on seed such as their quality, quantity, seasonal availability and areas of their availability are also collected periodically. Information on feed required for each species is also collected for this purpose. Data on economics of fishing operations in capture fisheries and culture operations are also collected to study the viability and capture and culture fishery operations. In addition quinquennial census on fishermen population covering their educational status, craft, gear availability and social aspects are conducted.

6. Data Analysis on Tuber Crops

K.H. Lakshmi

Central Tuber Crops Research Institute, Trivandrum - 695017

Much has been done for improving and increasing the productivity of tropical tuber crops. However, the role of climatic factors on the yield of tuber crops is yet to be quantified. It has been observed that meteorological parameters play an important role in the economical yield of various tuber crops. Parameters like total sunshine hours per day, rainfall, temperature and number of rainy days at crucial growth phase affect the yield potential of the tropical tuber crops. Elaborate studies are yet to be carried out to investigate and quantify the association of these triats on yield of tuber crops. Data already available on these aspects need to be analysed to provide additional information. Some such studies carried out are indicated.

Data generated from the project on "Cassava based multiple cropping system" by Ghosh et al. 1978 & 1989 and Kabeerthumma et al. 1985 were analysed to find the suitability of tuber crops as intercrop with perennials like banana, coconut, eucalyptus and Leucaena. Data on runoff and soil loss were also processed to infer the appropriate spacing and placing of plants to reduce soil loss. The cost and returns analyses indicated that cassava as a monocrop yielded the highest returns followed by cassava with intercrops groundnut and cowpea. Inclusion of cassava with banana and coconut improved the net returns substantially whereas with Leucaena and Eucalyptus the increase was modest.

Analysis of data generated from the trials of the project "Tuber crop based farming system for livestock feed" revealed that the yield advantages and recovery of nutritive components from total biomass indicate that certain crop combinations were more productive than sole cropping under the agro-climatic conditions of tropical South India under rainfed conditions.

Analysis of data generated from trails on fertiliser, spacing and uptake studies on tuber crops, have helped in finalising the package of practices for improved cultivation of all the tuber crops.

7. Exploratory Data Analysis Approach for Studying Retention Times in Diary Cattle

V.K. Bhatia

Indian Agricultural Statistics Research Institute, New Delhi 110012.

The culling of animals is a complex phenomenon and depends upon large number of factors operating on it. The net effect of culling thus results exceptionally long or short retention times of animals with different proportion of exotic inheritance. The presence of these types of extremes many a times disturb our inferences based on classical procedures such as mean and analysis of variance etc.

Keeping in view these practical problems of data analysis and interpretation of results thereon, the detailed careful examination of data through exploratory data analysis approach has been followed for studying the retention times of different categories of dairy animals. The data collected by IASRI on breeding and production records of different crossbred animals at the Ambala Military Dairy Farm have been utilised for the study. The examination of data has been carried out rigorously by using graphical and analytical procedures. Through graphical procedures such as histograms, stem-and-leaf diagrams, box-plots, normal plots, spread-level plots etc., the data have been thoroughly examined for extremes, points of concentration and gaps.

On the use of very simple analytical tools such as minimum, maximum, range, interquartile range, mean, median etc., the retention times are studied. It is observed that the lengths of useful life for animals with proportion of exotic inheritance greater than 25% are more variable in comparison to animals with less than 25% exotic inheritance. On examining the estimates of mean it is seen that the longest useful life at herd is for the crossbreds with 12.5% exotic inheritance and whereas the shortest is for the crossbreds with 87.5% exotic inheritance. But effectively the longest stay is of the crossbred with 37.5% exotic inheritance because they mature early and complete maximum number of lactations. Thus keeping in view of the complex definition of retention times, the average alone is not a proper measure for comparing the stayability of different crossbreds because it is highly influenced by very low and high values. It is clear that for the crossbreds 12.5% and 18.75% the distribution of stayability is symmetric and normal and because of this reason the estimates of location either based on mean, median, 5% trimmed mean or any robust estimator, the values do not vary much. On the other hand the values of robust estimates for the non-normal and non-symmetric distribution differ quite significantly.

Examining the estimates it is realised that there is a need to redefine the retention time of dairy cattle in the herd. Further on account of the presence of extremes, the robust estimates of location are also obtained and it is shown that it is quite different from the conventional measure of central tendency i.e. the mean. Based on robust estimates, the retention times of different categories of animals are being compared. Further on fitting of resistant lines, the points of little importance have also been identified.

From the fitted relationship equation, residuals are also worked out for exploring the influence of data points on the equation of relationship. It is quite clear that the absolute residuals for crossbred 6.25% obtained by fitting resistant lines to various location estimates are highest as compared to other crossbreds. Thus, due to wild residuals this implies that this category of crossbreds is having little or no effect on the line which summarizes the relationship. The other crossbreds whose residuals are also of significance are 25%, 50%, 75% and

87.5%. The only crossbred whose residual is smallest in all the situations is that of 37.5%. The study also identifies crossbred 6.25% as outlier as this influences little to the line of resistant fit. The residuals for this category of animals are quite large as compared to other categories of animals whether we take mean or median or Huber or Tukey's estimate of location. The second category of animals whose residuals are also sufficiently large is that of 50% exotic inheritance.

Finally it is concluded that when it is expected to encounter extremes, it is always desirable to use exploratory data analysis approach to draw correct and true inferences.

8. Use of Available Information in Agricultural Surveys

A.K. Srivastava
IASRI, New Delhi-110012

Sample surveys are considered a dependable and viable tool for generating data based information which are so essential for proper policy formulation of any sector of economy. Main emphasis in sample survey approach has been on descriptive uses where attempts are made to estimate certain population parameters like population mean or total or variances etc.

There is a growing awareness that in the usual sample surveys from various sources the collected data has got much more potential than what is harnessed from it. Thus, various data sources need to be utilised further for a better utilisation of available resources. Fortunately, the deeper analysis of such data is now possible through availability of computing facilities.

In the presentation, this point is illustrated in the context of small area estimation where it is tried to utilise the data collected in usual crop estimation surveys for developing estimates of average yield at small area level.

Synthetic estimation as a small area estimation approach

Small areas are basically small domains of the entire population. They are not necessarily geographical domains but they may consist of smaller parts based on ethnic, economic, social or age wise groups of individuals. Many a times such small domains are of immense importance. Sample surveys are normally planned for either entire population or for larger part of it and estimates for small areas need to be developed subsequently.

Several small area estimation approaches are available in the literature. However, synthetic method of estimation is discussed in detail in this paper. One of the assumptions in synthetic method is that auxiliary information in the individual cells is known and used as cell weights. In this paper an approach for estimating the cell weights, when they are not known, is suggested and the effect of this estimation is investigated. The auggested approach is illustrated through an example

Illustration

The approach of small area estimation is described on the basis of real data. The illustration is based on estimates of crop yields for wheat at block level in Gurgaon district of Haryana. The results are based on data for 1987-88.

In crop estimation surveys production of a crop is estimated as a product of average yield and area. Area figures are normally available on complete enumeration basis. While average yield is estimated through crop cutting approach. In Gurgaon district which consists of 8 blocks 138 crop cutting experiments were conducted. In the selected fields for crop cutting information on various inputs like fertilisers, seed rates, irrigation, etc. were also available. Important inputs were first identified through regression analysis approach. These important variables alongwith the yield rates were then used for making homogenous groups through cluster analysis approach. For this purpose SPSS package was used and minimum distance criterion for clustering data was used in order to make three homogenous groups. Previous 2 years data was used to augment the sample sizes within each group for estimating proportion of area under each group marginals. These group marginals were further used for making the estimates of cell weights, which are area under the crop in each cell. These weights were then used for estimating the crop yields for each of the 8 blocks. Estimates alongwith corresponding standard errors are worked out for synthetic estimator and those for usual sample estimator for 8 blocks in Gurgaon district. It is observed that the block level estimates for crop yield are fairly reasonable.

The illustration here provides an idea of this approach for utilising the available information collected in the process of crop estimation surveys. The information used here was otherwise not used in the crop yield estimation. As a caution, it may be mentioned that the approach needs to be verified further on the basis of some more experiments. However, the study indicates the potential of using small area estimation utilising available information for developing estimates of crop yields at smaller levels.