## Proceedings of the Session on
## Statistical Computing

The Indian Society of Agricultural Statistics organised a session on Statistical Computing during its 48th annual Conference at KAU, Trichur on 17 December, 1994, to highlight the role of Statistical Computing in Agricultural research particularly in analysis of data with sophisticated statistical techniques.

*Chairman*      :      Dr. S. Pillai
                       Joint Director (Retd.)
                       IASRI, New Delhi.

*Convenor*      :      Dr. V.K. Bhatia
                       Senior Scientist
                       IASRI, New Delhi.

At the outset, Dr. O.P. Kathuria welcomed the Chairman and highlighted the importance of this session particularly in the light of development of Statistical Computing techniques at a faster rate.

In his opening remarks, the Chairman summarised the status of Statistical Computing facilities available to agriculture and animal scientists. He also elaborated on the development of computer hardware technology and its capacities since inception to present state of affairs. In the end, he pointed out that the statistical use of computers is far below from its potential use and relating this he also focussed his attention on some of possible factors responsible for it, like non availability of suitable data bases in scientific research organisations, inadequate and improper data collection and storing equipments, lack of funds, lack of proper training and motivation to statisticians for the development of proper statistical software, no interaction between industry and teaching and research institutions and financial incentives etc. After this opening remarks, the Chairman invited Dr. C.V. Ramachandran Nair, Joint Director, Animal Husbandry, Kerala to present his paper. Dr. Ramachandran presented his paper entitled, "NICNET based Integrated Information system for animal health". In this paper he emphasised that for an overall improvement in animal husbandry and to maintain good health of the animals, it is important to have all the information pertaining to animal health in time. This information, if available in time, not only helps in complete eradication of diseases but also useful for the epidemiological and economic forecast for the future outbreaks of diseases. During his presentation, he highlighted the role of National Informatics Centre and discussed in detail the timely reporting of animal health status in the state through various state sponsored health schemes.

After this presentation, the Chairman himself presented his paper entitled, "Statistical Work and Computers". The main gist of his presentation was that how statistics should be taught by using computers extensively. The notion of central limit theorem, simple random sampling, unbiasedness, normality assumptions and many more concepts can easily be better understood by the use of computers. He also emphasised that many complex real life agricultural and biological phenomenon can now be visualised by using computers through the techniques of Monte Carlo Simulation. He was also of the opinion that efforts may be made for development of user friendly Statistical Packages so that users could use them without much of the training and could also manipulate the analysis as per the requirement. He was also of the view that the Statistical Package should contain the modern needs of the data analysis such as parallel processing, transformation and text editions etc.

After the presentation, participants had discussions, regarding the use of computers in teaching statistics, timely availability of relevant information not only for animal health but in general for the entire agriculture sector and about the availability of information highways etc. The following recommendations emerged.

–   Computers should be used extensively in teaching basic concepts of statistics;

–   User friendly statistical packages be developed for larger section of agricultural and animal scientists; and

–   Data bases be developed in different spheres of agriculture for timely reporting of vital information.

The detailed summaries of the two papers presented are as follows.

### 1.   NICNET Based Integrated Information System for Animal Health

*M.M. Ismail and C.N. Ramachandran Nair\**
Senior Systems Analyst, National Informatics Centre,
Planning Commission, Kerala State Unit

The contribution of Livestock to the National Economy is significant and the gross value of output from the Livestock Sector alone, at the current price was about Rs. 588 billions in 1992-93 which is about 26% of the value of the total agricultural output. This excludes the contribution from animal draught power. About two thirds of the animal protein consumption is derived from

\*   Joint Director (LP) & Nodal Officer for Computerisation, Animal Husbandry Department, Kerala State

milk, meat and eggs. Hence the management of this vital sector deserves considerable attention. The widely distributed livestock is likely to get outbreak, natural disaster etc. Thus this unique resource poses a very difficult situation for not having the right type of information to conserve and manage it.

There is a need for basic information to lay down scientific norms for the control and eradication of animal diseases. This would enable reliable assessment of the incidence and trends of the diseases. It would also help in making epidemiological forecasts and in planning disease control programmes more rationally and economically.

In the traditional system, only limited data regarding the prevalence of animal diseases were collected. In the system of reporting, the types of diseases and the details of animals affected and related information furnished by the veterinarians at the grass root level were compiled at the District level and then sent to the State level, where it was again compiled and finally sent to the National Headquarters.

The collection, compilation and analysis of data were all time consuming with the result that the money, energy and time spent on this particular exercise were wasted. Further there was a significant time lag in the flow and availability of information at different levels and this affected the quality of decisions. Therefore the information flow should be such that not only the information becomes available in time, but also the quality and reliability are ensured.

Based on the problems faced by the Management, a NICNET based Information System for the monitoring and management of animal health activities in the State has been designed and suitable software 'DISNIC-MISAH' (MIS for Animal Health & Administration) has been developed and implemented by the National Informatics Centre, Kerala State Unit in association with the Department of Animal Husbandry, Kerala. This package is used to process the data on health status of various species of livestock and morbidity pattern obtained from institutions located in various parts of the State on monthly/adhoc basis (within a week of reporting). The data resources relating to environment like climate, rainfall, atmospheric temperature, seasonal fluctuations etc can be made use of in analysing the animal health situation so as to take corrective and controlling decisions at the appropriate level, including epidemiological forecasts.

The Kerala State Animal Husbandry Department switched over to the Computer Based Information System from April, 1993 and so far it has been found that not only it reduces the burden of bulky reports of field level

functionaries but also the animal health status of the State is within the easy reach of the senior executives and policy makers.

The administrative and financial aspects, for which the Veterinarians have to make a report have also been included as functional modules in the Software. The reports generated by the system provides feed back to the field level functionaries.

This system thus helps in co-ordinating various information streams and provides interlinkages between sub systems, bringing them into an objective and integrated information system and will improve the quality of information, adding value to it for effective discharge of functions of personnel at operational and supervisory level. This will also enable the country, if all the States take up the system in a uniform manner, to supply reliable data in time to International Agencies, like the FAO, OIE, WHO which consolidate world wide statistics in respect of animal diseases.

## 2. Statistical Work and Computers

*Dr. S.S. Pillai*
Jt. Director (Retd), IASRI, New Delhi.

Statistical analysis of data collected in various investigations has become more and more detailed and refined with the advent of computers. The classical techniques used for data analysis can now be modified to suit the specific characteristics of the problem under investigation, as assumptions made for derivation of generalised theory may not be valid in all cases. The analysis can be done through the use of software specially prepared for the purpose or through ready made packages. This paper gives a general comparison of these two methods. It also discusses how statistical concepts can be appreciated by the beginner though practical exercises on computers rather than through strict mathematical proofs. The theoretical approach often makes the students to concentrate on the rigour of mathematics and they do not get the practical implications. This makes them incapable of using the concepts studied in a practical situation unless the training is followed up with analytical work under an experienced statistician.

Numerical verification of the concepts embedded in Central Limit Theorem, analysis of variance, testing of hypothesis etc. require that we should be in a position to generate random variables which follow specific distribution functions with known parameters.

One reliable method given by Knuth D.E. to generate random numbers on a computer is by using the Linear Congruential method.

Using the ability to generate random variables following any distribution function many of the basic ideas in statistics can be brought home to the students by practical exercises on the computer.

*Statistical Packages.*

A number of statistical packages are available on mainframe computers and PCs. These packages can be used for statistical analysis in a number of situations. A comparative study of all packages with their strengths and limitations appeared in one of the issues of BYTE in the mid· eighties. One of the popular packages used in PCs is SPSS/PC+, a review of which appeared in Nov 86 issue of BYTE (1986). The package provides full screen editor for preparation of data files as well as the statement file. It also provides for on line error prompting while statement file is prepared. Most important features of SPSS/PC+ are its ability to permit transformations of datavalues, grouping and recoding of data, computation of wide range of descriptive and inferential statistics, preparation of multiway tables of frequencies, correlation and regression analysis and handling of missing values.

The major drawback of all statistical packages is that they require training on the part of the user to handle them. An agricultural scientist who has conducted an experiment cannot get it easily analysed on a computer even if he has access to it without the help of an experienced hand. In short, the packages are not user friendly, as they should be.

One of the methods adopted by developers of software packages is to build into the product the features of a programming language so that it can produce object modules to help users to specific jobs without any training. An example is the use of computers in departmental stores for preparation of bills for items sold to a customer. A simple procedure is possible because programming packages like DBASE III are capable of producing object codes with linkages to files containing item code and its descriptive name and the unit price as applicable on the day. The software is capable of keeping tables and updating them independent of the billing clerk. If the statistical packages have similar object module preparing capability, simple user friendly routines can be produced for each type of analysis. For example, if the agricultural scientist can interactively enter the number of replications and treatments in a experiment conducted with ranumised Block design in response to prompts followed by data items for each replication and treatment, the rest of the analysis could be done by the computer. For using such routines no special training is required. There is need for developing such a statistical package with programming language capabilities.

There are a number of statistical applications which are not covered by the packages. The generalised routines for analysis of variance do not, for instance, help in analysing data collected in factorial experiments. Analysis of survey sampling data is also not easy except for simple surveys like opinion polls and pre poll enquiries. When stratified multistage designs are adopted with pps sampling at one or two stages or stratification also in one or more stages packages are not much useful. If they can be used at all a lot of ingenuity will be required to prepare the statements for the analysis. Moreover the execution time and storage overheads may be very high.

In packages transformations and groupings are done in a single pass for all cases when a statement is interpreted and executed. This makes the analysis very slow when the cases to be processed are large. A user written program can on the other hand process all the required analysis in a single pass of the data file for a number of tables. For large scale surveys and censuses, user written software should be preferred.

A general survey sampling analysis program can be prepared by devising menu driven inputs for different stages, getting different population parameters and sample size at each stage and then computing the multiplying factor (raising factor) as relevant to the ultimate stage unit which is the data item read one by one foraggregation. The parameters may be known at design stage or can be computed during the final run of the edit programs.