Validity and Reliability of Statistical Procedures[1]

G.R. Seth
Former Director, IASRI, New Delhi

I am glad to be with you in celebrating the 50th anniversary of the Indian Society of Agricultural Statistics. It is almost the same period for which I have been in contact with the subject of Agricultural Statistics. I started as a statistician in 1945 when I joined the Statistical Section of the ICAR headed by Dr. P.V. Sukhatme. As a matter of fact the history of the development of Agricultural Statistics in India is almost that of the Society and that of the Statistical Section of the ICAR. The latter is now Indian Agricultural Statistics Research Institute, IASRI. We sadly miss the presence of Late Dr. V.G. Panse who together with Dr. P.V. Sukhatme was the founder and builder of the Society and IASRI. If Prof. P.C. Mahalanobis, founder of the Indian Statistical Institute is considered the father of development of Statistics in India, then Dr. P.V. Sukhatme and Late Dr. V.G. Panse are rightly considered as the fathers of development of Agricultural Statistics in India. Thus on the golden anniversary of the Society we remember Dr. V.G. Panse and pay our homage to him and welcome Dr. P.V. Sukhatme and wish him a long life to continue guiding the Society in its work. I myself feel greatly honoured for being invited to deliver Dr. Panse Memorial Lecture on this occasion. It will not be out of place if I say a few words about the development of IASRI for which Dr. Panse spent the best years of his life.

ICAR in its right wisdom created a Statistical Section in 1936, to help it in the statistical examination of the agricultural research projects to be sanctioned by ICAR for their suitability and for the statistical evaluation of the results achieved as well as for the task of improvement of Agricultural Statistics. Dr. P.V. Sukhatme joined as Chief of the Statistical Section in 1939-40. He was able to gather round him a number of young enthusiastic capable persons to help him in this task. The Statistical Section produced results which were greatly appreciated by the ICAR. One of the outstanding achievements was the development of crop-cutting surveys for estimating yield of crops in India. The system so developed, I understand, is being adopted as a matter of routine by most of the States in India.

---

1 Dr. V.G. Panse Memorial Lecture delivered at the 50th Annual Conference of ISAS held at Vigyan Bhawan, New Delhi on 19 December 1996.

latent resources which are lying unutilized in the economy and can unleash wider processes of agricultural growth and decentralized developments linked to national and world markets. The perspective plan in the Ninth Plan may therefore amongst other strategies have to:

(i) institutionalize the Accelerated Irrigation Benefit Programme for completing large and medium irrigation programme;

(ii) operationalize the strategy of perspective planning for the water sector, being worked out by the special Commission set up for this sector;

(iii) raise fixed investment in the agricultural sector to at least 12% of agricultural GDP; work out the private investment inducing aspects of public investment in the water, land development, agricultural technology and research and agricultural markets sector;

(iv) carry forward the agro-climatic planning strategy;

(v) complete the process of agricultural economic reform, including tarrif reform, removal of trade bottlenecks, rural credit reform and liberalization of the cooperative sector; and

(vi) work out the institutional support base for small farmers and employment for landless labourers.

The good work done by the statistical section was recognised by ICAR when it upgraded the post of the Chief of the Statistical Section almost to the level of two current Commissioners - one for agriculture and the other for animal husbandry. Latter were the main advisors to ICAR for its agricultural research in the country. Dr. P.V. Sukhatme thus became virtually the third Commissioner in the form of Statistical Advisor. He helped the ICAR in close statistical scrutiny of the research projects submitted for its sanction. At the ICAR regional meetings he came in contact with the Directors of Agriculture and Animal Husbandry and agricultural statisticians to know about agricultural research in the states and their statistical requirements. This greatly helped him in orienting the statistical research and training programme at IASRI to meet those needs. Dr. P.V. Sukhatme resigned from the post of Statistical Advisor in 1950-51 to join the FAO as Head of its Statistical Division and Dr. V.G. Panse took over the post of Statistical Advisor. Dr. Sukhatme helped IASRI by encouraging FAO to use the facilities of the IASRI for the training of the personnel of the developing countries in Agricultural Statistics for carrying out the agricultural censuses and for the development of their agricultural statistical systems. FAO also provided services of famous statisticians to IASRI to tone up its statistical research and training programmes. Besides expanding its statistical research, it considerably expanded its training facilities. It started short courses for the States' personnel and for persons deputed by foreign countries in South East Asia, Arab countries and Africa. This training was amply utilised by these countries. During my visits abroad I met a number of trained persons. They admired and appreciated the training they received at IASRI. IASRI also organised a number of International Centres in Agricultural Statistics and agricultural censuses both in India and abroad. IASRI attracted good students for its M.Sc. and Ph.D. courses in Statistics which were started in collaboration with IARI when the latter received the status of the Deemed University. Also many foreign research workers worked at IASRI for short periods. Indeed it was a golden era for IASRI. All this helped the Institute to gain a good standing in the statistical world. This status could not have been achieved by IASRI but for the foresight, wisdom and hard work of Dr. V.G. Panse. Even though he had no formal training in statistical theory, he had an excellent sense of sound statistical procedures needed for the conduct of sample surveys and conduct of investigations in the agricultural fields.

In appreciation of the good work done by the Statistical Section, ICAR provided a building of its own in 1954 and unofficially gave it the status of the Institute in 1959. Formal recognition followed in 1968 after the re-organisation of the ICAR and it received its independent status in 1974 with the declaration of its status as an Institute, one expected its research and training programmes to get a further boost but unfortunately the reverse happened. It lost much ground in its research and training programmes. It is difficult to pinpoint the reasons for this setback. I can only offer a few surmises. After the re-organisation of the ICAR, it ignored the role of Statistics in agricultural research when the post of Statistical Advisor was abolished and a junior post of Assistant Director General (Statistics and Economics) was created in its place. In fact, DDG (Agriculture or A.H.) assisted by ADG (Statistics and Economics) took over the work of the statistical affairs of ICAR. Services of IASRI were no more utilised by ICAR to meet its statistical needs. As a consequence Institute lost its contact with the agricultural research of the ICAR as well as with that of the states. Many senior statisticians serving on the government side left the Institute to retain their status in the Indian Statistical Service and there was no equivalent replacements of the staff. FAO also gradually reduced their support to the Institute. The flow of good students from the states for its M.Sc. and Ph.D. courses also diminished with the introduction of these courses in the State Universities. ICAR also reduced its training programme. Perhaps the review committees set up by ICAR to help the Institute in adopting its programme to the changed situation did not make workable suggestions to revitalise its activities. I am sure that ICAR and its present Director are concerned about the Institute's continuing setback. It is high time that Director and the workers at the Institute should do something on their own. They should put their heads together to redress the situation and find ways to improve the working of the Institute. Director has to provide the leadership and workers should not ask what the Institute can do for them but tell the Director what they can do for the Institute.

Let me now turn to the topic of my lecture i.e. validity and reliability of statistical procedures. I will confine myself to only finite populations which are studied by sample survey statisticians. There is an explosion of tech-information in the world. Misinformation forms a part of it. It is difficult to distinguish "false" from the "true". Some of the users have abused the science of statistics by either culling of the data, or by using fictitious data or by employing improper statistical procedures to gain support for their views about certain public issues or to support their findings. This does not mean that statistical science is at fault. It only means that it has been mis-used. But the layman does not understand the validity and reliability of statistical methods. As a result the science of statistics gets a bad name. It is sometimes the worst of the less when they say-lies, damned lies and statistics. It is thus necessary for all of us to know about the validity and reliability of statistical procedures. Before the 20th century, survey investigations were usually carried out on a complete enumeration basis. Sometimes a part of the population called the sample, also formed the basis of an investigation. But later, it was realised that considerable time and resources were being spent in collecting data and analysing it by adopting complete enumeration of the population. A question was raised whether one can get satisfactory results by studying only a part of the population to get reliable estimates. However, then sampling could not be adopted as earlier sample studies did not convince its usefulness for wider adoption. In the earlier studies purposive sampling was adopted which really amount to quota sampling. Such samples were questioned for their representativeness of the population under study. Such samples could at best represent population in respect of one or two characters for which quotas were fixed. Such samples could not represent the population for which quotas were not fixed. Also the representation of the sample for which quotas were fixed, was also considered doubtful as quotas were not fixed on any objective basis but on the advice of an expert. Efforts were made to find some other means of getting representative samples. Suggested criteria only begged the question instead of solving the problem. Such criteria could not be used in practice as these were usually also the objectives of the study. Kiaer was probably the first person who successfully applied the method of random sampling in an enquiry in 1895.

It was Prof. Neyman who gave a sound foundation to the use of random sampling for the study of finite populations. He showed that quota sampling cannot give a representative sample for a finite population. Also it looks impossible to choose a criterion for selecting a sample which can reproduce the value of a parameter of the population. This objective can be achieved only by complete enumeration. In practice, however, even a complete enumeration may fail to provide the exact value, when the frame representating the sample is incomplete. Thus the question of representativeness is to be considered from some other angle. According to Prof. Neyman one can develop a sampling method which is capable of producing estimates, which are as close to the exact value of the parameter as desired, but only with a high probability. This probability can be as high as 95% to 99% but never 100%. Thus in certain cases, we may find that the estimates of parameters are not that close to the parametric value. He showed that random samples chosen with a suitable probability of selection can do this. He used Gauss-Markoff's theorem to prove that the best linear unbiased minimum variance estimate of the mean value of any character of a population exists and its value will be as close to the exact value as desired, but only with a high probability. Samples which can give estimates of the population parameters with known sampling errors will be called valid samples. Random sampling can produce such samples. These sampling errors can be controlled with proper choice of a scheme for probability system. So even a simple random sample, where equal chance is given to all possible samples, will give estimates with known sampling error and thus is a valid sample, but it will have a larger sampling error and thus its reliability will be low. This reliability can be raised by choosing a better probability scheme. By reliability, I mean a degree of closeness of the estimate to the parametric value. The closer it is, it is considered more reliable. The question of validity also arises regarding the procedure adopted for estimation of a parameter from a valid sample. In case estimation procedure does not follow the probability scheme for sampling, the estimation procedure is invalid.

By reliability we mean the reliability of the estimates produced by a valid sample. Invalid samples will lead only to invalid estimates. By the better reliability of an estimate we mean the following. Estimate 'A' is considered more reliable as compared to an estimate 'B' of the same parameter if 'A' is closer to the value of the parameter than 'B' with a high probability. Rigorously speaking, an estimating procedure 'A' is more reliable than an estimating procedure 'B' if estimates obtained by adopting procedure 'A' have lower sampling errors or shorter confidence intervals for a given probability of their inclusion in the confidence interval. Efforts of a statistician are to obtain estimates of high reliability for the given resources at his disposal.

Sampling statisticians have done an excellent job of producing better estimates to meet the needs of the users. They, however, have to face criticism of theoretical statisticians that they are not adequately utilising the efficient statistical techniques developed for infinite population size. There is some truth in their allegation. However, in any field, advances in theory take a lot of time to be adopted in the applied aspects of that field and sample surveys are no exception. It may, however, be said that theoretical statisticians have not yet come up with estimates which are better than those currently in use by the sampling statisticians. Also there is a psychological barrier in the use of infinite populations for preparing estimates for a finite population. Finite population is to be considered as a part of a super-population. Thus, we are replacing an actual population by an imaginary infinite population. Thus, we lose our grip over the actual population in hand while searching for a suitable imaginary infinite population to produce better estimates. Capability of a prior super-population to produce robust estimates has also not yet been established. Also the users will have less confidence in estimates based on a super-population. What I am trying to say is that the use of a super population to produce better estimates is not that easy as it is made out to be.

Bayesians too criticise the work of sample survey statisticians. They say that randomness and sampling designs for drawing a sample are not necessary for estimating population parameters. Bayesians, as you know, make considerable use of a priori distributions in their work and they recommend the same approach to be adopted by sample survey statisticians. Prof. Savage, a well-known Bayesian himself does not agree with the suggestion of giving up altogether the use of randomness in selecting the sample and advocates some use of it when he says that some randomness will be always necessary to produce valid estimates. I agree with that view. One should study the role of "controlled selection" and "balanced samples" in the formulation of better designs and use as much a priori information available in similar surveys carried out in the past for drawing the sample. But use a suitable probability scheme for selecting the ultimate units. Thus efficient probability designs should continue to be developed while controlling the selection and measurement biases.

There is another problem raised by Dr. Godambe regarding the efficiency of the estimates for finite populations. Prof. Neyman had used the Gauss-Markoff Theorem to develop the best linear unbiased estimate for a finite population. Dr. Godambe made a fundamental contribution to the sampling theory by indicating that application of Gauss-Markoff theorem for developing best linear unbiased estimates was not strictly applicable for finite populations. Following up the general estimate given by Thompson and Horvitz, he deduced that more general type of linear estimates were possible for finite populations and he proved that in the more general linear estimates, a uniformly minimum variance unbiased estimate does not exist. While asserting this he missed the distinction between a "finite" and an "infinite" population concerning the preparation of estimates. For an infinite population the serial number of a selected unit is the only consideration for forming estimates whereas for the finite populations, there are other criteria also available besides the serial number of the unit. The selected units in a finite population are real. They have an identity or label and these can be identified. Also a probability sampling design like stratification or selection of units with probability proportional to size can be adopted for selection which is not possible in the case of infinite populations. Prof. Tikkiwal developed seven class of linear estimates of which T1-class is the one which considers only the serial number of a selected unit for forming a linear estimate. This T1-class coincides with the class of linear estimates used for infinite populations. Prof. Neyman had used this T1-class only for the preparation of linear unbiased estimates. Thus use of Gauss-Markoff theorem is justified in this case. One can have a minimum variance unbiased estimator in this class. It has also been proved that in the T1- class there is a minimum variance unbiased estimate: Prof. Kempthorne looks at the better estimate given by Dr. Godambe to illustrate the non-existence of the best linear unbiased estimate. According to Prof. Kempthorne the estimate provided by Dr. Godambe was sterile from a practical angle. The proof given by Dr. Godambe requires the knowledge of an inequality between the parameters of the population. In practice one would rarely know of this relationship. If one had that knowledge, then it can be utilised to find a better estimate than the conventional estimate. Sarndal supports this view when he says if the actual is known to belong to a known set of the parameter space, one should try to find out estimators better than the mean. I am not trying to minimise the importance of Dr. Godambe's fundamental of non-existence of uniformly minimum variance unbiased estimate in the general class of linear estimates. What I am trying to say is that no better usable unbiased estimates have been so far developed than those presently available. Sampling statisticians should continue to find better usable estimates than the mean even if a uniformly better estimate does not exist for the general linear class of linear estimates.

Now let us consider the ways in which the validity of sample can be affected during the execution of a survey. Besides the sampling error, errors are committed in the execution of the survey. Such errors are called non-sampling errors. Much attention has been paid for reduction of sampling errors while designing a sample survey. Comparatively, little attention has been paid to the control of the quality of data to be collected in a survey. Non-sampling errors can occur at any stage of the survey - from selection of units to the final stage of preparation of the report. These errors will occur even when the survey is done by complete enumeration. In fact contribution of non-sampling errors is much more in complete enumeration as these errors have a tendency to rise with the increase in sample size. Sometimes contribution of these errors to the total error is so much that estimates so prepared become totally useless and unreliable for any decision making process. When this happens, then there is a total loss of resources spent on the survey. Statistician in-charge of the survey is no doubt conscious of the possibility of these errors escalating at each stage of the survey. He tries all steps to control them. However, the execution of the survey at each stage involves the play of human and other factors over which he has no control. Frame used for sampling may be incomplete. It is required to be updated during the field work. The enumerators in such a case are asked to list any units noticed by them which were not listed in the frame. Enumerators are usually asked to select a sub-sample for the non-listed units in the frame according to instructions regarding the method of selection as given by the statistician in-charge. This method of selection should not be left to the enumerator. If done so, different enumerators will adopt their own methods of selection. The result will be that their inclusion in the sample will make the sample invalid for the probability of selection of the realised sample will become unknown. As already mentioned, different factors introduce their own type of errors. Such errors have a cumulative effect in the sense that these errors affect the quality of field work in its later stages. Thus, it becomes important to know how to control errors as they arise in the execution of the field work. Attention can then be paid more to the errors which are likely to severely affect the reliability of the data under collection. We have already discussed the incompleteness of the frame used for selection of the units. Next is the question of non-response. It can be total non-response of a selected respondent. The respondent is no more there in the area or it refuses to respond. Enumerators will then be asked to substitute fresh units for enumeration and proceed further as they have been instructed to do for the incomplete frame. Units substituted should be indicated in the filled-in proforma. In case of a partial response, where the respondent is unable to give the information on certain items of the questionnaire if the enumerator himself can provide the information if he can do so. However, it should be indicated on the proforma which information has been so supplied.

It is for the statistician to decide how to use the information supplied by the enumerator. Besides the above errors mentioned, there will be the question of the quality of response. The latter is indeed difficult to evaluate. Errors can be reduced by clear-cut instructions as well as by trial of questionnaires in the field as mentioned earlier. Enumerators and supervisors will make unintentional as well as intentional errors in the field. The unintentional errors can be due to lack of proper understanding of the instructions, lack of co-operation of the respondent and play of nature which prevent them from working in their full. The intentional errors will be their lack of interest in the field work, absentism, filling the questionnaires without doing the field work, etc. Supervisors may be slack in their supervision, casual in their supervisory work etc. It is indeed difficult to enumerate all the causes which are likely to affect the quality of field work. In the end, I can only say that the quality of field work depends heavily on enumerators and the supervisory staff. In case they are honest and devoted workers, they will find ways and means to overcome the problems faced in the field and produce data of high quality. Thus a heavy responsibility falls on their shoulders for the success of the survey. Data collected is the only basis for the analysis of data and production of the report.

Lastly, I will like to talk more about the control of measurement errors. There are various ways in which the response is measured. The choice of the method or methods will depend upon the planning strategy of the statistician and the resources available for carrying out the field work. This includes the quality of enumerators especially their experience of field work. I mention usually the following methods which are used in the field for measuring responses (1) mailed questionnaires (2) interviewing respondents (3) direct observations like eye estimates without measurement (4) objective measurements (5) registration of events like births and deaths (6) Institutional records like those kept by patwaris for land use or records kept by customs department. Wherever relevant, use of records is recommended. Data from these records can also be used as a check on the information collected by other methods. There is a possibility of their use in providing auxiliary variables for improving the estimates of some other survey where the characters under study are highly correlated with the characters noted in the records. Birth and death records will be very useful in population studies. Mailed questionnaires approach is probably the cheapest method of collecting the information especially in opinion surveys. But there will be lot of non-response. Also the quality of data may not be as good as the survey done by interviewing the respondents. But the latter is costlier compared to that of mailed questionnaires. Perhaps a suitable combination of these two may provide more reliable results. I did not mention the survey by contacting persons only on the telephone. This is really suitable only for 'yes' or 'no' type of surveys. But these are highly biased as they exclude respondents who do not have the facility of a telephone. Regarding non-response, measures suggested earlier shall have to be taken to restore the validity of the sample. Objective measurements approach is the best method of recording data. This considerably reduces non-response and produces the best quality of data. But it is very costly. If cost is not the consideration this method should be adopted. To reduce the cost, it can be combined with a larger sample of eye estimates of which the measured units will be a sub-sample. Objective measurements can be measured in various ways and the statistician should choose the method of measurement which is likely to give more reliable estimates. For example, crop cutting experiments indicate that a bigger plot compared to a small circular cut gives better estimates. Latter has a tendency to include more crop on the circumference of the circle compared to a large rectangular plot. Thus it overestimates the yield.

It will not be possible for me to give you here examples of the many studies carried out in India by the various organisations like Indian Statistical Institute, IASRI, etc. In India we have made much progress in developing methodology of estimating crops yields, cost of production, animal husbandry products like milk, wool and fish catch. Studies like fish catch involve complicated design with the sampling of sites, days and hours. In all those studies emphasis has always been laid on the objective measurement, clear-cut instructions to the enumerators and intensive supervision of the field work.

Our experience is that it is impossible to totally eliminate non-sampling errors. But it is the duty of the institution incharge of the survey to provide some idea of the accuracy of the results. Also successively over time, the data should have less non-sampling errors.

Indian Statistical Institute and IASRI have suggested two methods for checking as well as improving the field work. Adoption of these methods may also give some idea about the non-sampling errors occurring in the field work. Indian Statistical Institute suggests that interpenetrating samples should be used, each sample to be conducted by a different set of enumerators so that one can have a number of estimates of the characters under study. If the various estimates so obtained are compatible one can consider that the field work was satisfactory and one can combine these estimates to get a better estimate. However, in case they widely differ, it will indicate that quality of work is not good and it will be difficult to produce a reliable estimate. IASRI suggests the method of rationalised supervision. This method involves splitting the sample entrusted into two random parts (1) to be canvassed by the enumerator and (2) the other to be done by enumerator-cum-supervisor. IASRI claims that the introduction of supervisory checks will yield better results. Supervisor spends more time with the enumerator and as he understands the instructions to do the work in a better way, enumerator is encouraged to do better work in his unsupervised work. Also he will become a better enumerator for future surveys. In case the estimates from the two samples-one from the "enumerator-cum-supervisor" subsample and the other from the "enumerator's work" subsample agree, they will be combined to provide the estimates. In case the two sub-samples provide incompatible results, the sub-sample worked by "enumerator-cum-supervisor" team is expected to be more reliable and will provide the estimates of the characters under study. Also comparisons of the raw data of the two subsamples may provide an idea why and where the subsample canvassed by enumerators failed to provide reliable information. Thus the method suggested by IASRI is likely to yield better results. It also perhaps provides some suggestions for better conduct of field work in future.

In conclusion, I submit that randomness has worked well in practice and has produced valid samples to provide reliable estimates. Sampling statisticians are doing a good job of developing better and better estimates. There is always a scope of improvement. They should pay more attention to the suggestions made by theoretical Statisticians and Bayesians. Perhaps better use of a-priori information is recommended. In survey reports I will like more information on the quality of work and how the quality of work can be improved in future surveys.