

STATISTICAL ERROS AND BIASES – A REFERENCE TO MANAGEMENT SCIENCES*

A.Z. Memon¹

I will be mainly speaking about the statistical errors within the context of statistical applications made in various management sciences.

The major functions of a management science are planning, organizing, staffing, leading and controlling. This science whether dealing with administration, human resource, strategic planning, marketing, finance, production, inventory etc., it uses quantitative for one purpose of the other function. For instance:

- A manager administration in an organization may have to analyze the feedback data about his employees
- A human resource manager may have to bank on statistically oriented criteria for the selection of staff.
- A product designer calls for statistical assessment and analysis of consumers needs. He needs this information to design a product or for its modification in connection with taste, design and performance.
- A marketing manager may have to evolve or compare advertisement strategies for the promotion of a product.
- A financial investor has to do the capital budgeting and create optimal portfolios using linear programming techniques.
- A strategic planner has to base his plans on statistically forecasts.
- A production manger has to follow statistical quality procedures to produce quality products, or to determine the probability of its breakdown during specified period. Inventory management uses calculus to decide the optimal order size for the replenishment of inventory.

¹ Institute of Management Sciences, Lahore

*A talk given at the statistics Conference, G.C. University, Lahore in October 2003

certain scientific principles. The element of reliability is a part and parcel of this function and is an essential requirement.

The other main function is that he studies numerical data to facilitate knowledge that serves as a basis, a support, a service for various decisions in sciences, industry, business and government.

The need for numerical information to the management has always been felt in the past as a source to entail or guide decisions for good or evil purpose, for constructive or destructive aims. Even in ancient times, crude methods were employed to determine a population's parameters. To quote an ancient data collection technique I refer to Herodotus, a Greek historian. According to him, when Emperor Xerxes wanted to make a decision to attack Greece, he was worried. He was known to be worried because he had no knowledge regarding the strength of his army both for its management and invasion. The story goes he or his advisors conceived an ingenious idea. He assembled a myriad of 10000 soldiers in a close huddle, drew a line around this group, raised a low wall on the line. This measuring pen was emptied and filled 170 times to estimate puzzling manpower parameter of that time. Those who were involved must have been presumably duty bound to collect correct information for fear of handing sword over their heads. But all that was a minor problem of counting, measuring a certain statistical parameter.

The scope of statistical function has kept on increasing even when there has been a loud clamour that there are lies, damn lies, and statistics.

The expectation from a statistician has always been that he increases the amount and relevance of information in an inquiry. In institutions or places where the statistical data are poor the policies, the decisions, their political, social, economic dynamics and developmental programs based on such data may not be only fruitless but could be harmful as well. In medical research the harm is visible before it is long, but the results of a developmental policy may become manifest after a long time. Poor information leads to contradictions, conflicts, disorder, chaos, or confusion. It produces false signals and misleading illusions. A leader is expected to have right information if he is to set up right goals, build up right vision, and lead his team towards this vision. A manager is also expected to have right information if he is to do things rightly.

POOR STATISTICAL INFORMATION

The possible causes of poor statistical information lies in the problems.

- i. Surveys/experimental plans, questionnaires may be poorly designed, incorrect application of statistical techniques is used. Faulty interpretation of statistical results.

- ii. The number of units/items/respondents may be inadequate for inferential purposes.
- iii. Biases occur in measurement/collection analysis and interpretation of data.

The first problem may be attributed to a number of causes.

- Ignorance of relevant statistical knowledge

It is important to note that Statistics is a body of tools. Each tool has a specific function, its own assumptions, its own limitations. It is possible that two or more tools perform the same function. Statistical expertise lies in choosing a tool that best befits the parameters of a situation. The efficacy of information largely depends a lot on achieving a greater coordination between theory and practical limitations. Given a situation one tool can do a lot better than the other, but errors can not be always avoidable.

Such errors are probable everywhere. One very serious error was made during American election in 1936 where in an attempt to predict the vote for President Roosevelt, Gallop underestimated more than 19%. This figure came as a shock for many statisticians. It was discovered latter that the population taken was representative of only a certain cross section of the society.

- Lack of statistical advice, guidance or consultancy

This problem can be attributed to the limitation of resources or material. It is something beyond one's control – like in medical where the number of patients of a rare disease is very small. An observation of medical nature could be very expensive as well. Whatever the case, a statistician should inform his client about the inferential repercussions when the number of observations is small.

- In sufficient observations for inference

This problem can be attributed to the limitation of resources of material. It is something beyond one's control – like in medical where material. It is something beyond one's control – like in medical where the number of patients of a rare disease is very small. An observation of medical nature could be very expensive as well. Whatever the case, a statistician should inform his client about the inferential repercussions when the number of observations is small.

- Lack of interest or awareness about the importance of statistical inquiry.

It is again a factor that adversely affects the quality of information. The research shows that in a society where the importance is well realized, well recognized, well appreciated the quality of the response from its members is significantly high. Our data collecting agencies work very hard to collect information, and attempt very hard to collect accurate data but it is the respondents that also matter.

The quality of information essentially suffers when biases occur. Faulty measuring equipment is one source. Poor questions also introduce distortions. Printing errors are also possible. Let us ignore errors that are not intentional or deliberate. This problem may be caused by a statistician, or his client, or field staff or respondents.

CLASSES OF STATISTICIANS

Statisticians are generally divided into three broad in view of ethical considerations. Statistical theorists, subject specialists (who have dual knowledge of statistics and another science – genetics, medicine, economics, business etc), fact finding statisticians.

FACT FINDING STATISTICIAN

The third group include consultants for statistical operations. This group also includes those involved in management of statistical inquiries. They have to come in contact with the clients who hire them, staff that assist them in conducting inquiries, and respondents who are included in the inquiry. It is here that the problem of ethics could be a vital issue.

EXPECTATIONS

The question is what are our expectations from practicing statisticians.

A statistician's job is to find the truth nothing but the truth in an inquiry, and to unfold it he is expected to be a person of proper knowledge, skills and experience, intellectually honest, patient, rational, logical, analytical. He should be a leader capable of taking thorough care of the survey work and giving it meticulous attention. Regarding his relation with client, if his client wants the findings of his study to stay confidential his trust should not be betrayed. Regarding his relation with respondents if a respondent does not want to be quoted, he should not be quoted or identified. Regarding his staff, he must train them fully and affectionately,

explaining the urgency and usefulness of their participation in the survey and the reasons as why to be intellectually honest (to avoid natural biases and non-sampling errors). He should be able to generate caution, care, devotion and seriousness in his staff regarding the conduct of their duties.

STATISTICIAN' PROBLEMS

But a statistician is also exposed to various problems from his clients, from his staff and respondents.

BEHAVIOUR OF CLIENTS

His clients may be individuals, firms, organizations, government that want a statistical inquiry to be conducted, or a data to be analyzed. All clients do not want objective inquiries. Clients could be nasty, and make unreasonable requests or exercise an improper pressure to pervert certain features of an inquiry to achieve preconceived notions, theories or interests. They set up certain targets for their statisticians to achieve through deceitful support by using data, even by concocting data. Let me quote the following from an international journal where a statistician writes.

"I became aware of this early in my medical consulting career when in a cooperative venture to organize some data for presentation in a legal case a physician suggested calculating the average survival time of a group of cancer patients using the data from "

A government statistician may also come across similar experiences. Furthermore, a statistician may also get threats from those whose interests are at risk because of his discoveries.

BEHAVIOUR OR FIELD STAFF

The field staff of a statistician gathers all the primary information. They could be honest, diligent, cooperative, dedicated, concerned, involved or otherwise. Their statistical conduct is important in preserving the purity of information. Without their participation a statistician may never have access to true information. There may as well be a reason for their abnormal attitude. I remember in Sierra Leone during my FAO assignment in Africa, a senior statistical enumerator in the agriculture department told me that he used to cook data whenever he was assigned rural areas in the country where cannibals were known to dwell. So it was primarily his fear which forced him to adopt a non-professional practice.

In Lesotho, once I was giving a lecture to field staff on the precision of length measuring equipment and its effect on the acreage estimates statistics at national level. I met there a field reporter who instead of actually measuring the length of a sample field would estimate the length by counting the number of paces his horse made through a field. At my inquiry why data from the back of his horse he said he was scared of snake bites. So it is understandable that it is some fear that induced abnormality in his behaviour. There may be a number of other reasons as well. Personal convenience is generally the main reason. At times an enumerator exploits the respondents in rural areas behaving as a powerful representative of the government.

RESPONDENTS

A respondent may be willing or unwilling, serious or non-serious, uninvolved, detached, indifferent, non-cooperative, contending in supplying information. He may have certain suspicions and fears - one is why me in the sample? He is scared of being exposed, humiliated, blackmailed, overtaxed. To him an inquiry may have no meaning, some thing senseless and he behaves irresponsibly, inconsistently or non-cooperatively. Experience show that this kind of his behaviour has much to do with the way he is approached by a statistician or his staff.

STATISTICIANS THEMSELVES

As already indicated the second function of a statistician is to provide information through surveys and data analysis. We must distinguish between poor statistical work and unethical statistical work. The differentiation factor is intent. A statistician could be unethical in many ways. An unethical statistician may in the following activities.

- He Manipulates Fake Data
 - He may find a false probability to show that a company's product will be acceptable in the market.
 - He may play with a (3 to falsely support the research hypothesis.
 - He may create a subjective investment portfolio.
 - He may create false perceptions when analyzing feedback data.
 - He removes outliers from data without any justification it. We have to remember that each observation is to be treated sacred and valuable unless there is a strong reason to see it the otherwise.

As a matter of fact he may manipulate any statistical information if he wants.

- He Engages In Plagiarism Of Data.
At times statisticians may copy or steal data from other sources
- He connives with the management or clients to produce deceptive information. For example, an unethical statistician falsely supports the effectiveness of a certain medical treatment against some disease and thus helps the client to dodge and exploit the innocent community.
- He sells expensive information regarding a client to its competitors to give them competitive edge. In designing market strategies, an organization often needs statistical information on predicted market shares. The concerned statistician may sell this valuable information to its competitors.
- He behaves indifferently to statistical requirements when designing surveys, preparing questionnaires, fitting statistical models, applying testing procedures, analyzing data, and interpreting results.
- He Blackmails Both Clients And Respondents
Let me be a little more specific with reference to the following few statistical issues of ethical interest

SURVEYS

A survey should be objectively informative, rationally expressive and professionally reliable. Ethical considerations arise with respect to generally four types of potential errors when designing surveys.

- Selection Bias becomes an ethical issue when particular individuals are excluded from the population to favour the sponsor or other person.
- Non-Response becomes an ethical issue when particular individuals are less likely to be available to respond and the statistician knowingly designs the survey to exclude them.

- **Sampling Error** becomes an ethical issue when the findings are intentionally presented without reference to sample size to promote a viewpoint that might otherwise be truly insignificant.

A bigger standard error gives a larger confidence interval (CI), and is likely to include the stated null hypothesis (H_0), and so lead to a non-significant result. If this SE is not mentioned, the sample estimate appears significant.

- **Measurement Error** becomes an ethical issue when an interviewer through his mannerism and tone guides the responses in a particular direction.

PRESENTATION OF RESULTS:

- One complaint against a statistician is that he sometimes distorts the visual impression to create false message when displaying diagrams and charts. For instance, the vertical axis may be compressed to relate two variables.
- Statistical ethics demand that both good and bad results arising from a research deliberation should be documented. Bad results are equally useful in research.
- Unethical behaviour occurs when a statistician willfully goes for an unsuitable measure (the mean or midrange for a very skewed data) to distort the reality to support a particular position.
- The failure to include a confidence interval estimate may mislead the user of the survey results into thinking the point estimate from the sample is all that is needed to predict the population characteristic

DATA ANALYSIS

Most of statistical techniques are based on assumptions. Without verification of those assumptions the statistical results could be misleading. Statistical ethics demand that this aspect of analysis should not be ignored.

For example, with reference to multiple regression a statistician has to look into the following aspects:

- i. Assumptions (independence of errors, constant variance, normality of observations)
- ii. Selection of variables (relevant variables to be considered)
- iii. Multicollinearity (which if not removed it distorts the influence of regression coefficients)
- iv. Treatment of outliers

Unethical behaviour occurs when someone willfully ignores the above issues.

Furthermore data snooping (that is looking at the results and then deciding whether to select one tailed or two tailed test, or choose the level of significance) is never permissible in the world of statistical ethics. Similarly, meta analysis (combing and summarizing previous research results) is a controversial issue.

CONCLUSION

I have discussed some vital aspects of statistical errors. I am sure practicing statisticians who remain deeply involved in statistical work may have a lot to add more. I stop here with a quotation from the Holy Quran which includes the essence of what may directly pertain to statistical ethics and that is - Do not clothe the truth with falsehood and conceal it when you know it.