ERRORS IN RESEARCH DESIGN

A useful way of looking at the marketing research process is that involves the management of errors. At all stages, from problem formulation to report presentation – errors can arise. It is rare that a research project will be error free. Consequently the research designer must adopt a strategy for managing and maintaining the error. We shall first look at the component of errors and then the type of errors.

COMPONENTS OF ERRORS

The objective underlying any research project is to provide information that is to be accurate and error free as possible. Maximising accuracy requires that "total error" be minimised. Total error has two distinct components:

Total Error = Sampling Error + Non Sampling Error

Total error is usually measured as total error variance – also know as the mean squared error:

$(Total\ Error)^2 = (Sampling\ Error)^2 + (Non\ Sampling\ Error)^2$

Sampling Error refers to the variable error resulting from chance selection of elements from population as per the sampling plan. Since it introduces random variability into the precision with which sample statistics are calculated it is also called random sampling error. **Non Sampling Error** consists of all other error associated with the research project. Such errors are diverse in nature. They are often thought of as *bias*. However bias is a type of systematic error which enters into the process because of un-calibrated instruments or prejudices of the researcher. There can be completely random components of non-sampling error. For example a misrecording of a response during data collection represents a random non-sampling error. However using deliberate "loaded" question to get desired response – which maintaining the status quo of randomness – is called bias. Thus non-sampling errors can arise from various sources – but in marketing research project they arise mainly from response and non-response.

To a large extent the major error components are related to each other in an inverse manner. Increasing the sample size to reduce the sampling error can lead to an increase in non-sampling error. In most statistical projects – the effort is to minimise the total error by minimising the sampling error and assuming that good training of the investigators will reduce the non-sampling error. However in market research that may not be the possible as the time and cost constraints may prevent choosing a sample size which will allow us to be confident that the sampling error (and hence total error) has been minimised. Ideally a market researcher should strive to minimise both the types of error. Considering the time and cost limitations this can rarely be done. The market researcher must make a decision that involves a trade off between sampling and non-sampling errors. Unfortunately very little is known about the relative size of two error components – although some empirical evidence suggests that non-sampling errors tend to be larger of the two components. Sampling errors can to a large extent be reduced and/or controlled by following statistical procedures, but such a check is generally not possible on non-sampling errors

TYPES OF ERRORS:

1. **Population Specification Error:** It is defined as the "non correspondence of the required population to the population selected by the investigator". It occurs when a researcher selects an inappropriate population from which to obtain data. For example many times package goods manufacturers conduct surveys amongst housewives – because they are easy to contact and because it is assumed that as they are the end users they make the purchase decisions. This assumption may not be always valid since husbands and children may significantly influence the buying decisions.

2. **Sampling Error:** It is defined as the "non correspondence of sample selected by probability means and the representative sample selected by the researcher". It occurs when a probability sampling method is used to select a sample and this sample is not representative of the population concerned. For example, a random sample of 500 people composed only of people between 35-50 years of age may not be representative of adult population. Sampling error is affected by the homogeneity of the population under study. In general – more homogenous the population; the smaller the sampling error. Sampling error falls to zero in case of a census.

3. **Selection Error**: It is defined as the "Non correspondence of sample selected by non probability means and the representative sample sought by the researcher." In other words – it is a sampling error for the sample selected by non-probability methods. There is a natural tendency for the investigator to select those respondents who are most accessible and agreeable. Such samples are often comprised of friends and associates who more or less represent the desired population. Samples of this type may not be representative of the population but are certainly not selected randomly – leading to statistical problems in inferencing.

4. **Frame Error:** It is defined as the non-correspondence of the sought sample with the required sample. A *sampling frame* is a means of accounting for all the elements of the population. It is usually a listing of all the elements that identify a population. A perfect frame identifies each population only once and does not include elements of other population, which may lead to overlap. For example a sampling frame that specifies the population as all adults using oral care products – may well leave out people who prefer using *neem* and *babool* sticks, homemade toothpaste and powders – leading to a frame error in defining the population.

5. **Non Response Error:** A non-response error occurs, when the obtained sample differs from the original selected sample. Non response can occur in two ways – (a) Non-Contact i.e. the inability to

contact all the members of the sample and (b) Refusal i.e. the non-response of some or all the items of the measurement instrument.

Non-contact errors arise due to the inability to reach the respondent. This may be because the respondent is NAH (Not at Home) on the initial call (or call backups) or may have moved away from the area either permanently or for the period of the survey. Non Contact errors can be reduced by careful analysis of the selected sample.

Refusal arises when the respondent does not respond to a particular item or to multiple items of the questionnaire. Monthly Household Income, Religion, Sex and Politics are topics that may elicit item refusal – normally categorised as refused. Don't Know / Can't Say (DKCS) refusals arise when the respondent is aware of the item but is unable to provide an adequate answer to the item. Constraints of time on the part of the respondent, a bad experience with some previous interviewer, ill health etc. may be other reason for refusal. Refusal rates could be brought down to manageable levels by giving training to the interviewers and continuously monitoring the investigation process.

6. **Surrogate Information Error:** This is defined as the non-correspondence of the information being sought by the researcher and that is required to solve the problem. Or in other words, information is obtained from substitutes rather than original sample. The necessity to accept surrogate information arises from either the inability or the unwillingness of the respondent to provide the needed information.

   Decisional oriented behavioural research is always concentrated with prediction of behaviour usually non-verbal. This limits most marketing research projects to use proxy information – data from past behaviour. Attitudes, beliefs and SEC classification are all examples of surrogate information because based on these information we try to predict the future behaviour of the respondents. Secondary sources of data are another source of surrogate information. Surrogate information error can be minimised by ensuring that the information used is highly correlated with the actual information obtained.

7. **Measurement Error:** This may be defined as the non-correspondence of information obtained by measurement process and the information sought by the researcher. It is generated by the measurement process itself and represents the difference between information generated and information wanted by the researcher. Such errors can potentially arise at any stage of the measurement process – from the development of the instrument till the analysis of findings.

   The error can also occur at transmittal stage - when the interviewer is questioning the respondent. Faulty wording of question, non-preparation of non-verbal clues, behaviour of the interviewer etc may all contribute to how the respondent interprets the question.

   In the response phase – when the respondent is replying – error may occur because the respondent gave a wrong answer or the correct answer was wrongly interpreted and recorded. In the analysis phase, errors of incorrect editing, coding and/or descriptive summarisation and inferences can contribute substantially to measurement error.

8. **Experimental Error:** It can be defined as the non-correspondence of the "true impact of" and the "impact attributed to" the independent variable. When an experiment is conducted, the researcher attempts to measure the impact of one or more manipulated independent variable on some dependent variable – while controlling the impact of exogenous variable. Unfortunately control over all the exogenous variable is not always possible – which may lead to an inclusion of bias in the experiments. For example a study – trying to find out reasons for cancer may conclude that as all the patients under observation used to eat bread therefore eating bread is the cause of cancer.

CONCLUSION

For any research project, recognising that potential error exists is one thing – but doing something about it quite another matter.
There are two basic approaches for reducing errors.

The first is to minimise errors through research design. In this process effective use of research methods and techniques are utilised to lessen the impact of both sampling and non-sampling errors. However, cost constraints, and at times the peculiar nature of error they prevent complete minimisation of error through this method.

The second is to estimate and measure error. In spite of all the precautions undertaken, not all errors – especially those related to fieldwork – would be eliminated. In such a situation if we can have an estimate of error we can say how accurate the research design was. However only sampling errors are measurable with some degree of confidence.

Either way, estimating is not an easy task due to the peculiar nature of the errors. Statistics help us to reduce the sampling error to a large degree but for non-sampling error researchers still have to rely on their intuition.