Chapter

Data Collection

Data collection is a crucial step in all scientific enquiries. It forms foundation of the work.

2.1 MISTAKES AND ERRORS

Mistakes are faults which can be avoided and minimized, while the errors are the faults which can be minimized only. Mistakes are the faults committed due to: (a) lack of skill, (b) faulty observations, (c) wrong recording and (d) incorrect statistical calculations. These could be intra-observer or inter-observer. Errors are the shortcomings which occur even after careful observations made by a skilled and trained person. The errors are of three type:

- i. Instrumental error/technical error: These are introduced as a result of faulty and unstandardized instruments, improper calibration, substandard chemicals, etc. For example, weight of a person can be measured to the accuracy of 1.0 kg, 0.5 kg, 0.1 kg and 0.01 kg. Depending on the accuracy of the instrument even after using the weighing machine mentioned, at the end (i.e. the one weighing to the accuracy of 0.01 kg) some error in weight will exist because theoretically it is possible to weigh the person with more accuracy than the one with the weighing machine used. It is, therefore, rightly said that, all measurements are approximations.
- ii. *Systematic error*: This is a repetitive error introduced due to a peculiar fault in the

machine or technique. This gives rise to same error again and again. For example, if the 'zero' of the weighing machine is not adjusted properly, it will give rise to a systematic error. If dilution of chemical is erred, it will give rise to the error of same quantum in titrations.

iii. *Random error*: This is introduced by changes in the conditions in which the observations are made or measurements are taken. For example, a person may stand in different positions at two different times, when his height is being taken. The error due to this phenomenon will neither be constant nor be systematic.

Errors and mistakes can be minimized by following measures:

- i. Using standard, calibrated instruments.
- ii. Using standardized, pre-tested questionnaire.
- iii. Using trained, skilled persons.
- iv. Using multiple observations and averaging them.
- v. Using correct recording procedures.
- vi. Applying standard and widely accepted statistical manipulations.

2.2 RECORDING OF THE INFORMATION

The collected information must be recorded properly. Mistake is introduced if correct information is recorded incorrectly. Following are some of the precautions taken to ensure correct recording.

- i. Limiting the observations/observer ratio to a feasible/practical level.
- ii. Keeping writing/recording work to minimum.
- iii. Using standard units of measurement.
- iv. Good working conditions.
- v. Adequate supervision.

2.3 STORAGE

The information/data collected may be required at some time in future. So, adequate care must be taken to store the collected information, so that, it can be retrieved and used at proper time. Following are some methods:

- i. Use of durable stationary (viz. cards instead of paper).
- ii. Storage with proper precautions against destruction by insects and rodents.
- iii. Transfer into computer.

2.4 PRETESTING

It always a good idea to check the data collection tool before the big show starts. Checking is done for validity and reproducibility. It is called pretesting. Most common tool that requires pretesting is questionnaire. Pretesting is done on small sample in conditions similar to the main project. Unforeseen problems may surface during the pretesting. These are corrected before the main project is implemented. Pretesting should not be confused with pilot. Pilot is miniature of main study. It is also done in environment similar to the main study, but with finalized tools of collecting information after pretesting. Pilot may be done to train the personnel involved or for gathering information about variables for determining the sample size.

2.5 DATA COLLECTION METHODS

- i. *Measurement*: In this method the required information is collected by actual measurement in the object, element or person. If we are interested in hemoglobin percentage of the individuals, we actually measure the hemoglobin levels by appropriate method. The measurement and actual enumeration generates primary data.
- ii. *Questionnaire*: Here, a standardized and pre-tested questionnaire is given/sent and the respondents are expected to give the information by answering it. The success of this method depends on the quality of questionnaire, the enthusiasm of the respondents and the ability of the respondents to give accurate and complete information. By this method, the information about a large number of attributes and variates can be collected in a short time.
- iii. *Interview*: This method can be used as a supplement to the questionnaire or can be used independently. Here the information is collected by face to face dialogue with the respondents. The success of this method depends on the rapport established between the interviewer and the respondent, ability of the interviewer to extract the required information from the respondent and the readiness of the respondent to part with the information.
- iv. *Records*: Sometimes the information required can be available in various records like census, survey records, hospital records, service records, etc. The utility of the information depends on its uniformity, completeness, standardization, accuracy and the reasons for which the information was recorded.

6