STA 111: DESCRIPTIVE STATISTICS

LECTURES 2

Topics: Data Collection Methods and Sampling Techniques

By: E. E. E. Akarawak

SCHOOL OF SCIENCE AND TECHNOLOGY
PAN-ATLANTIC UNIVERSITY

Lecture Outline

-Types, Sources & Method of Collection

-Population Vs Sample; Sampling and sampling methods

Expected Learning Outcome

Learning Outcomes:

- At the end of this lecture, the students should be able to:

i. Outline different sources of data

ii. Explain methods of data collection

iii. Explain various techniques of sampling

DATA COLLECTION METHODS

Statistical method for problem solving involves the following:

- **1. Develop a clear and concise description of the problem.**
- **2. Identify the important factors that may affect this problem or that may play** a role in its solution.
- **3. Propose a model for the problem, using scientific knowledge of the** phenomenon being studied. State any limitations or assumptions of the model.
- **4. Collect data to test or validate the tentative model or** conclusions made in steps 2 and 3.
- **5. Refine the model based on the observed data.**
- **6. Manipulate the model to assist in developing a solution to the problem.**
- **7. Conduct an appropriate experiment to confirm that the proposed solution to the problem is** both effective and efficient.
- **8. Draw conclusions or make recommendations based on the problem solution.**

DATA COLLECTION METHODS

- The collection of data is the first step of statistical investigation
- Because many aspects of real-life practice involve working with data, it is important for scientists to learn how to collect data
- An effective data-collection procedure can greatly simplify the analysis and lead to a valid conclusion on the population or process that is being studied

Sources of Data Collection

- In Statistics, there two major sources of data collection.
- The Primary source, eg., survey, observation, experiment. More reliable
- The Secondary source: internet, textbooks, records, archives, etc
- Data can be further classified according to their sources. Hence,
  - Primary data: The first-hand user collects such data from primary sources.
  - Secondary data: The user only collects already ‘collected’ and/or used data from secondary sources.

Sources of Data Collection

- Scientific data can be collected in generally 3 ways:
- Retrospective Study based on historical data
- Observational Study
- Designed Experiment

QUIZ:

- Which of the above is primary and which is secondary.

Retrospective Study

- A retrospective study uses either all or a sample of the historical process data from some period of time.
- A retrospective study takes advantage of previously collected, or historical, data. It has the advantage of minimizing the cost of collecting the data for the study.
- THIS IS SECONDARY SOURCE/DATA

Observational Study

- As the name implies, an observational study simply observes the process or population during a period of routine operation.
- Usually, the engineer interacts or disturbs the process only as much as is required to obtain data on the system, and often a special effort is made to collect data on variables that are not routinely recorded, if it is thought that such data might be useful.
- With proper planning, observational studies can ensure accurate, complete, and reliable data.
- However, these studies still often provide only limited information about specific relationships among the variables in the system.
- You can observe a process over time – Time Series

Designed Experiments

- The third way that engineering data are collected is with a designed experiment.
- In a designed experiment, the engineer makes deliberate or purposeful changes in controllable variables (called **factors**) of the system, observes the resulting system output, and then makes a decision or an inference about which variables are responsible for the changes that he or she observes in the output performance.

Survey

- Method of collecting data from all members of a sample by questioning techniques, including
  - Questionnaire
  - Interview
  - Opinion poles
- The units of study are humans. For opinions, perceptions, and so on
- Frequently used in social, medical, behavioural sciences and sciences

Problems of Data Collection

- The main problems that may be faced when collecting data are in the selection of appropriate collection methods and in the training of the staff involved.
- Other problems include Language barriers, Lack of adequate time, Expense, inadequately trained and experienced staff, etc

Summary and Conclusion

- Remember, DATA is the backbone of Statistics
- Collect them with care
- If not, ……
- Directly or indirectly, you may be affected

Review of Population and Sample

- Population refers to the total number of individuals or units being investigated.
- **Challenge:** Almost always impossible to study the entire population
- Sample is a part of the population selected for analysis. Must be selected with care;
  - Must be random
  - Must be representative
  - Must reflect characteristics of interest in population
  - Sample size should be analyzable

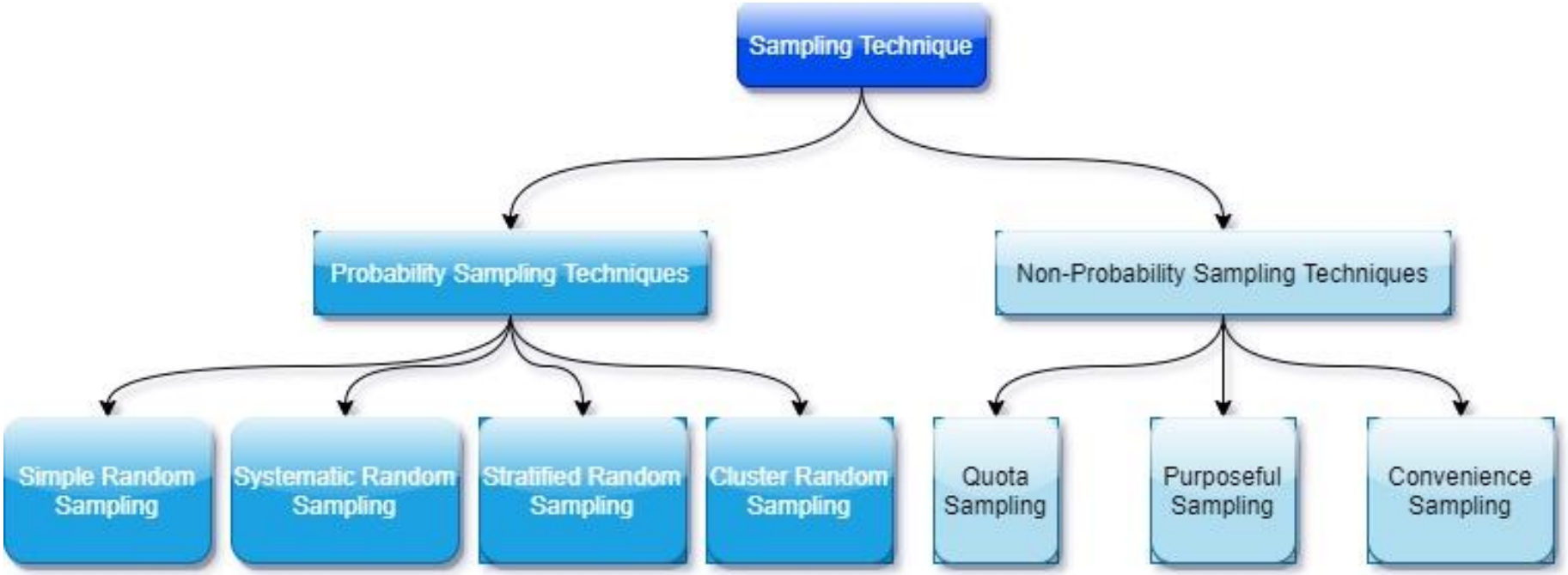
Sampling

- Sampling refers to the procedure for selecting a sample from a population.
- Sampling is very critical in statistical and empirical investigations
- Various sampling techniques are available depending on the goal of investigation

SAMPLING TECHNIQUES

- For inference purposes, a sample must be random and representative of the population.
- Sampling techniques can be split into two:
  - **Probability (Random) Sampling Techniques**: These refer to sampling methods which involve random processes. That is, the selection of the sample units is carried out by some impersonal (strictly determined) method and is uninfluenced by human choice.
  - **Non-Probability (Non-random) Sampling Techniques**: Any sampling method that does not involves any random process, but based on human judgement and feeling.

SAMPLING TECHNIQUES



Probability Sampling Techniques

- Simple random sampling
- Systematic random sampling
- Stratified random sampling
- Cluster random sampling

Simple Random sampling (srs)

- Allows every member of the population equal chance of being selected
- A sampling frame is required: Sampling frame is a list of all members of the population, together with their characteristics
- Example of srs techniques are: throwing a die, use of random numbers, balloting, casting lots, lottery, etc.
- A major challenge in using this method is getting a sampling frame.

Systematic Sampling Technique

- This is a sample in which every *k*th item in the sampling frame is selected after a random pick among the first *k* elements.
- For example, if we desire a 10% (*k* = 10) systematic sampling, we would determine the first member of the sample by using any random process to randomly select a number between 1 and 10. Suppose the random gives 6, then the first member to be picked would come from the 6^{th} position on the sampling frame. Then the second will be from the 16^{th} position, the third from the 26^{th} position and so on.
- *k* = N/n; where N = population size and n = sample size.

Stratified Sampling technique

- First, divide the entire population into distinct groups (strata) based on homogeneity or similarity. This procedure is referred to as Stratification
- Then, sample within each group
- Group size is important, to determine number to be selected from each group to make up required sample size, *n*.

Cluster Sampling

- Divide the entire population into distinct groups (called clusters) based on geographical proximity.
- The process is called clustering
- Then, sample within each group.
- Sample size allocation to group is similar to Stratified sampling.
- Sometimes, clustering can be in stages.

Non-Probability Sampling Techniques

- Refers to any sampling method that does not involve random process but based on individual judgement and feeling.
- Examples are:
  - Convenience Sampling: Sampling as is convenient
  - Purposeful sampling: Sampling just to achieve a particular purpose
  - Quota Sampling: Sampling to reflect different characteristics of groups in the population.
  - Snowball Sampling: Sampling mostly by referral

- Non-probability sampling methods are not reliable and prone to bias.