STA 111: DESCRIPTIVE STATISTICS

LECTURES NOTES FOR WEEK ONE

Topics: Introduction to Statistics

By: E. E. E. Akarawak

SCHOOL OF
SCIENCE AND
TECHNOLOGY
PAN-ATLANTIC UNIVERSITY

Before we get Started

- Familiarization
  - -My details: Dr. Akarawak, E. E. E.
  - -Your details
  - -With Statistics:
    - How would you define Statistics?
    - Do you use Statistics? How?

- Why Statistics?
  - -Uncertainties
  - -Variations

*What on earth?
*Who on earth?
*Where on earth

Lecture Outline

- Definition and Role of Statistics
- Definition of other Concepts
- Data and Data Classification

Expected Learning Outcome

Learning Outcomes:

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

i. Define Statistics as a course of study or body of knowledge

ii. State the role of Statistics in different fields of study and Engineering in particular

iii. Define some terms used in the study of Statistics

iv. Classify data

Definition of Statistics

Statistics, as a course of study or body of knowledge, can be defined as:

- A branch of science that deals with collection, presentation, analysis and interpretation of **data**.
- It is the scientific process for making valid decisions in the face of uncertainty.
- It is the Science of processing data.

Areas of Statistical Study

- There are two broad area of Statistics study;
- Descriptive Statistics and Inferential Statistics.
- Descriptive Statistics: Concerns itself with methods for presenting and summarizing sample data. Also referred to as exploratory data analysis (EDA)
- Inferential Statistics: Has to do with methods of using the summary information and findings from sample to draw conclusion on the population. Can lead to confirmatory data analysis (CDA).

Role of Statistics

- It is noteworthy that statistics can be used in everyday life. Its role in the life of modern society is enormous.
- Statistics put much importance on the analyses of data, which helps incorporate theory into solving problems of uncertainty. These theories inform the methods to help establish scientific foundations to problems and their solutions.
- Statistics play key roles in the state economy, by providing summary measures of economic variables and acts as a management tool.
- In Health, Energy, Environmental Studies, Government, Telecommunication, Transportation, etc, Statistical findings help to release the right information needed in policy formation and decision making.
- Statistics is increasingly used in risk assessment and dynamics.
- Statistics is also used in Control Theory and so it is essential for every scientist to master these tools.

Some Roles of Statistics in Software Engineering and Computer Science

- Probability and statistics are used throughout engineering to analyze data.
- Statistical methods are used in developing and implementing data-driven technologies.
- When data are generated in software cycle, statistical methods are use to describe, estimate and make predictions.
- Statistics methods provide frameworks that helps in identifying trends and patterns in data and these are useful in business decisions
- Data science techniques like machine learning and Artificial intelligence rely on statistical tools for analysing and implementing big data.
- Engineers and computer scientists use probability in product and system deigns.
- Combining Statistics knowledge and computer science creates increasing cutting-edge opportunities for many in the world of today.
- Engineers use probability and statistics to assess experimental data and control and improve processes.
- It is essential for today's engineer to master these tools.

Some Basic Concepts in Statistics

- Variable
- Experiment
- Population
- Sample
- Parameter
- Statistic

VARIABLE

- A **variable** is any characteristic, or quantity that can be measured or counted or observed.
- For example, hair color is a variable which can vary from person to person.
- Another example is a person's *height*. It is a variable which when measured would have different values for different people.

Types of Variables

- Data are collected on variables. So, understanding different types of variables can help one understand the data being studied.
- Qualitative variable: Values are not numbers. E.g. Gender, Programme of study (student to name some)
- Quantitative Variable: Values are numbers. E.g., Monthly income, production rate of a machine, lifespan of equipment, Energy use, etc

EXPERIMENT

- An **experiment** is any defined process that can yield results
- Random (Statistical) experiment is any experiment whose results cannot be influenced or pre-determined.
- The outcome of a statistical experiment depends on chance; the outcome can not be influenced or predicted.
- A common example of a statistical experiment is a coin toss. It's outcome as we all know is up to chance.
- Give an example of experiment that does not depend on chance (engage students)

POPULATION AND SAMPLE

- A **population** is the aggregate or totality of all the items or units under study (can be conceived of as the universal set).
- For example, measurements of the diameter of all possible bolts as they come out of a production process would make up a particular population.
- A **sample** is a chosen part of the population in question, say the measured diameters of twelve bolts chosen to be representative of all the bolts made under certain conditions. It is a subset of the population.

PARAMETER

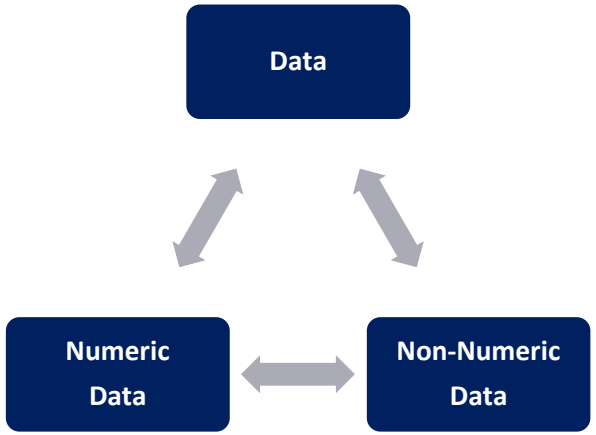
- In statistics, a parameter is any quantity of a statistical population that indexes or describes the population.
- If population is reachable and observable, the values of a parameter can be computed from population values.
- Examples of parameters are population mean, μ; standard deviation, σ; proportion, P; etc.
- Challenge: Population is not always reachable, so the values of parameters are often unknown (reasons)

STATISTIC

- This is a quantity whose numerical value is obtained from sample data.
- It is to a sample what **parameter** is to population.
- That is, the value of a parameter is calculated from the population values, whereas the value of a statistic is calculated from sample values.
- If for example the mean is calculated from a sample, then it is a statistic and not a parameter. Although it can be used as estimate of the population parameter.

Data

- DATA: Data are raw unprocessed information.
- There are basically 2 types of data:
- Qualitative (non-numeric) data
- Quantitative (numeric) data

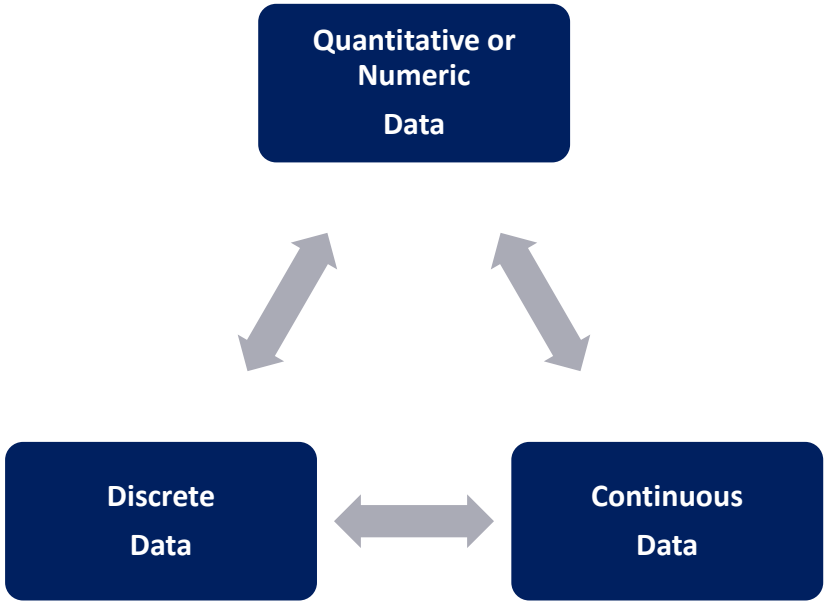


Quantitative Data

- Data which are numbers. Quantitative data is of two types:
- Discrete data: Countable whole numbers or integers. Examples include: number of people in a room, number of apples on a tree etc. Arise mostly from a counting process.
- Continuous data: numbers within an interval or uncountable range of numbers. Examples include: person's height, weight etc. Arise mostly from a measuring process.

Quantitative data (cont'd)

- Discrete data cannot be fractional values as it makes no sense to have half a human being, whereas continuous data can be fractional.

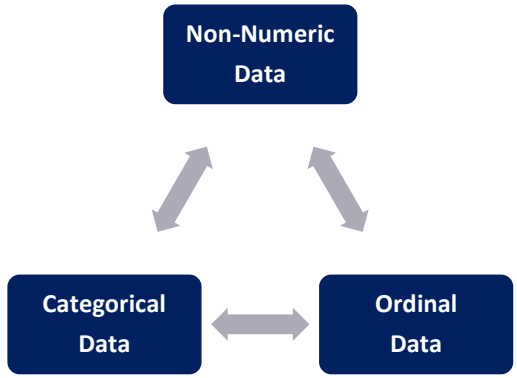


Qualitative Data

- Data which cannot be quantified. Examples include: hair color of people in a room, blood groups, DNA, RNA, country, gender etc.
- Qualitative data are **Ordinal** or **Categorical**.
- Categorical data can be divided into groups. Each group has no clear ordering. Examples:
  - Hair Color: Which can have categories *Blonde, Red, Brown* etc.

Qualitative Data *(cont'd)*

- Ordinal data can also be divided into groups. Here, the groups have some natural ordering to them. Examples:
  - The categories *Low, Medium, High,* have a natural order to them. Low < Medium < High, even though these are not numerical in nature.



Class work

Identify each of the following as examples of (i) Non-numeric (ii) Discrete (iii) Continuous variables:

- The hair colour of people in a concert show.
- The number of hours required to heal a patient of a disease.
- The length of time required answering a telephone call at a certain business center.
- The number of pages per job coming off a computer printer.
- The kind of trees used as Christmas tree.
- The number of voters in a community.
- Whether a statement is true or false.
- The number of books in a library.

Data Classification According to Measurement Scale

- Needed to determine the appropriateness to use certain statistical analyses
- Three important attributes that define a scale of measurement are: order, distance, absolute or true origin or absolute zero.
- Nominal data: has none of the attributes, eg, matric no., labels, etc.
- Ordinal data: has only order. Eg., low, high, very high. Mathematical operations make no sense

Data Classification According to Measurement Scale (*contd*)

- **Interval data:** has order and distance. Intervals between values are the same, difference makes some sense but not division. Eg., temperature, age group, chronological year.
- Ratio data: has order, distance, absolute zero. Division makes sense. Eg., length, income, voltage, etc. In age, someone who is 40 is twice as old as someone who is 20