Business Statistics

Objectives and Scope of Statistical Studies

> Statistics is the science of learning from data, involving collecting, presenting, analysing and interpreting data.

Objectives

- **Primary Objective:** Obtain information about a target population using a sample
- **Target Population:** Comprises all relevant subjects of interest
- **Sample:** A manageable subset, selected to make the study feasible. A sample answers questions about a target population.

Scope of Statistical Study

Follows the structure:

- **S**tudy Design
- **A**nalysis of Data
- **I**nterpretation of Results
- **D**ata moving forward

Study Design:

- Formulate the **question** of interest
  - What? Who? Why?
- Specify the **target population**
  - Who/What? Where? When?
- Determine the measurements to be collected (the **variables**)
- Define the **method** of data collection
  - How? When? Where?

Populations and Samples

- Target Population should be well defined
- Sample should be representative of the target population (not biased) and large enough to give accurate information about the population
- Ideally the observations should be independent of each other.

Selecting a Sample

- Only a **representative sample** should be used to make inferences about the target population.
- This is unbiased and large enough to give accurate information about the population
- One way to ensure that a sample is representative of the target population is to obtain a random sample
- A **simple random sample** is where each member of the population has the same chance of being selected

A Representative Sample:
Given the difficulty in obtaining a simple random sample, researchers must ensure they obtain a representative sample where the characteristics represent those of the target population without bias.

Types of Studies
Classifier:

1. **Deductive vs Inductive** (evidence based)
2. **Qualitative vs Quantitative** (nature of evidence)
3. **Observational vs Experimental** (interaction with variables)

➢ Deductive (non-empirical)→speculative
➢ Inductive (empirical)→involves collecting data and facts
  o Qualitative (unstructured)
  o Quantitative (structured)
    ▪ Observational: No intervention by the investigator
    ▪ Experimental: Control over the determinant

Variables:

- Measurements are taken on subjects in a study according to the variables of interest
- These measurements vary from one subject to another (possible predictors)
- Collecting data is the process of collecting values of the variables
- Values that differ randomly between subjects are called random variables
- Variables = quantifiable measures
- Values = quantities of variables
- Identifier Variables = Categorical variables with the purpose of assigning a unique identifier code to each individual in the data set (eg: phone number).

In any study, variables take on specific roles and these roles are classified as:

| predictors | influence | outcomes |

Where the predictor is affecting variables and the outcome is the affected variables.

Bias: Sample Size

Bias is any systematic error which results in an incorrect estimate of a parameter or an incorrect association between variables in study.

Selection Bias
Any systematic differences occurring in the way that subjects are selected for a study. Distorts representative quality.

Measurement Bias:
Systematic differences in the measurement of variables (methodology distorts sample data)

Response Bias:
Occurs when the response rate to a survey is to low. Should be at least 75%. The participation of sample distorts representative quality.
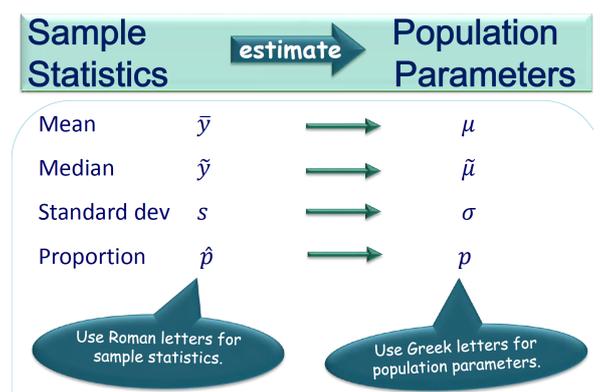
Confounding:
A confounder is a variable that distorts (increases or decreases) the apparent effect of one variable (determinant) on another (outcome). As such the correlation is confused as causation due to a third variable.

Eg: Suggested 4 hours of watching TV associated with heart disease. However, likely those aren't exercising enough which causes the risk.

A sample size needs to be sufficiently large to give an accurate representation of the target population. Accuracy of a sample for determining a population characteristic depends on 2 factors:

- *Sample Size* (n) used for the study
- *Variability* (spread) of the measurements

Sample Statistics		estimate →	Population Parameters
Mean	$\bar{y}$	→	μ
Median	$\tilde{y}$	→	$\tilde{\mu}$
Standard dev	s	→	σ
Proportion	$\hat{p}$	→	p

Use Roman letters for sample statistics.

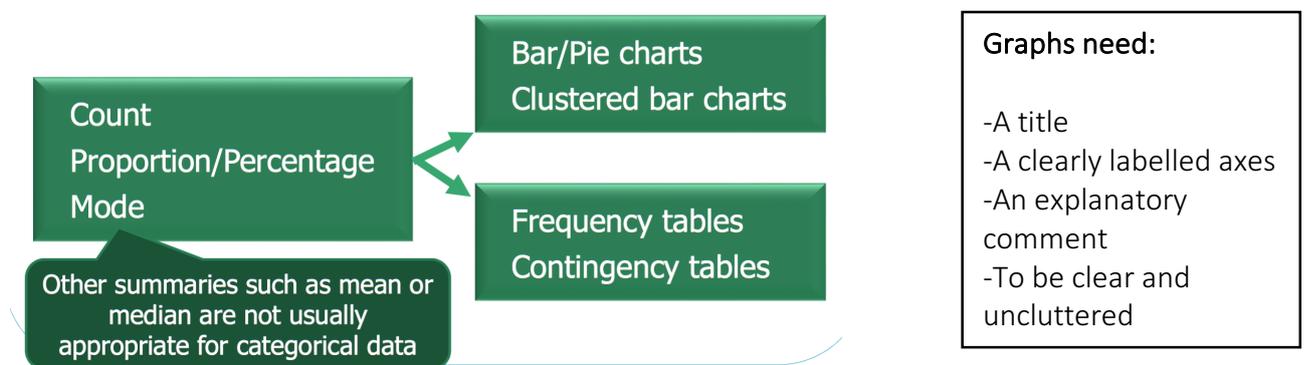Use Greek letters for population parameters.

Data Classification

Categorical Variables

Variables for which each observation falls into 1 of a number of groups. If there are 2 groups, variable may be referred to as binary or dichotomous.

-Nominal Variables have no inherent ordering
-Ordinal Variables are grouped with ordering

Summarising a Categorical Variable

Count
Proportion/Percentage
Mode

Other summaries such as mean or median are not usually appropriate for categorical data

Bar/Pie charts
Clustered bar charts

Frequency tables
Contingency tables

Graphs need:

-A title
-A clearly labelled axes
-An explanatory comment
-To be clear and uncluttered

1. Frequency Tables
Records the counts for each of the categories of the variable. The table should show the sample size and also:

- Variable name
- Name of each category
- The count, proportion and percentage of observations in each category

Bar Charts:
Obeys the area principle, giving an accurate visual impression of the distribution of a categorical variable, showing the counts for each category. Should have small spaces between bars to indicate that these are freestanding bars that could be rearranged into any order. Variable name is often a subtitle for horizontal axis (y=frequency)