Session 2:

Review of Basic Concepts in Statistics

What is Statistics?

- The science of collecting, analyzing and making inference from the collected data.
- It is called as science and it is a tool.

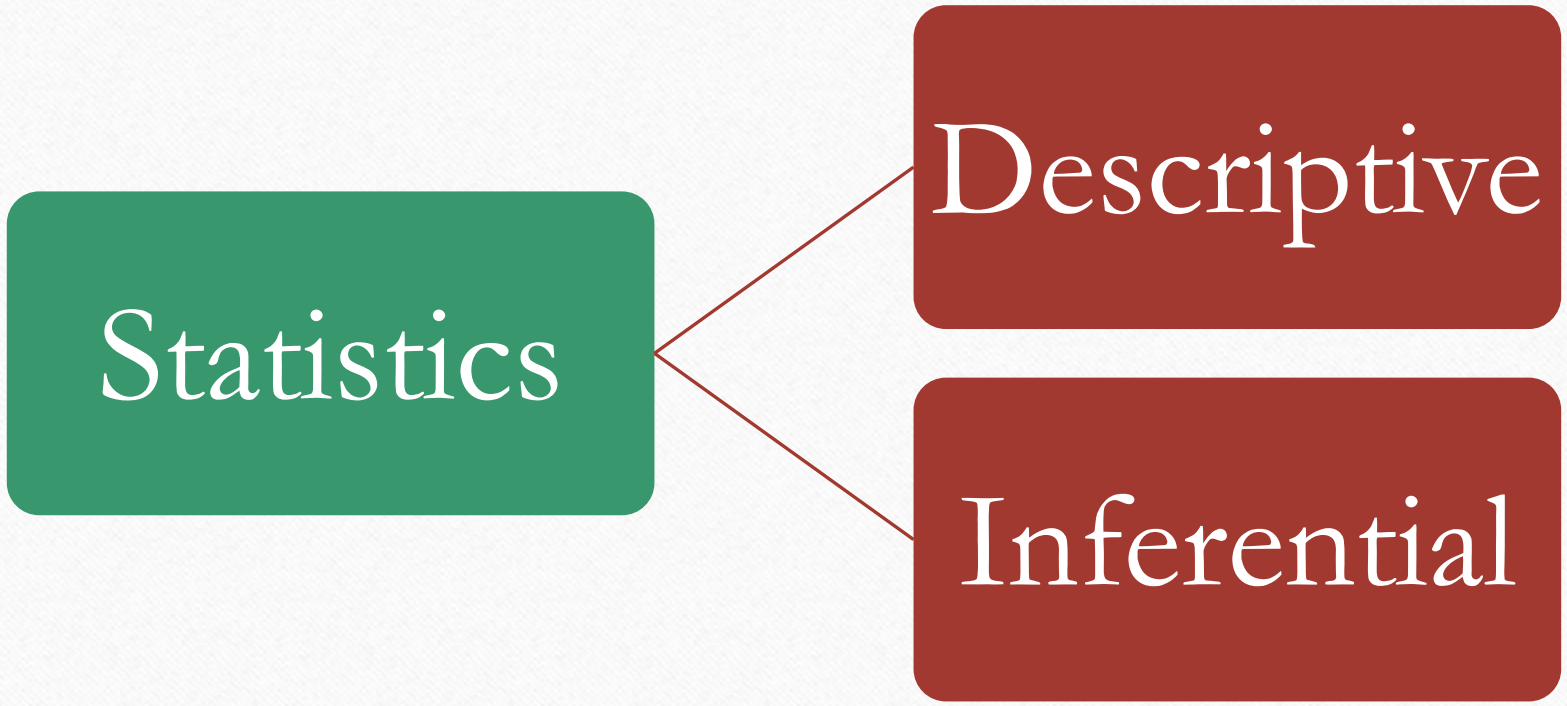
Statistic vs Statistics

- **Statistic**:
  - It means a measured (or) counted fact (or) piece of information stated as figure.
  - e.g., height of one person, birth of a baby, etc.,
- **Statistics**:
  - It is also called Data.
  - It is Plural.
  - Stated in more than one figures.
  - e.g., height of 2 persons, birth of 5 babies etc. They are collected from experiments, records, and surveys.

Why Statistics?

- Statistics is used in many fields:
  - Medical statistics
  - Agricultural statistics
  - Educational statistics
  - Mathematical statistics
  - And so on…

Types of Statistics



Descriptive vs Inferential

Descriptive Statistics:

- Once the data have been collected, we can organize and summaries in such a manner as to arrive at their orderly presentation and conclusion.
- This procedure can be called Descriptive Statistics.

Inferential Statistics:

- The number of birth and deaths in a state in a particular year.

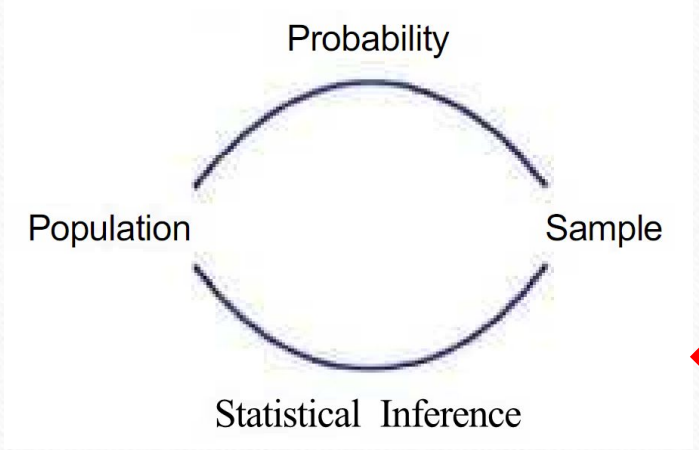
Sample vs Population

- Information is gathered in the form of samples, or collections of observations.
- Samples are collected from populations that are collections of all individuals or individual items of a particular type.

The Role of Probability

- Elements of probability allow us to quantify the strength or "confidence" in our conclusions.
- Major component that supplements statistical methods and help gauge the strength of the statistical inference.
- The discipline of probability provides the transition between descriptive statistics and inferential methods.

Probability vs Inferential Statistics



- For a statistical problem, the sample along with inferential statistics allows us to draw conclusions about the population, with inferential statistics making clear use of elements of probability.
- Problems in probability allow us to draw conclusions about characteristics of hypothetical data taken from the population based on known features of the population.

Sampling Procedures

1. Simple Random Sampling
2. Experimental Design

Simple Random Sampling

- Implies that any particular sample of a specified sample size has the same chance of being selected as any other sample of the same size.
- Sample size: the number of elements in the sample.
- Biased sample: A non-random sample of a population in which all elements are not equally balanced or objectively represented.

Experimental Design

- A set of treatments or treatment combinations becomes the populations to be studied or compared.

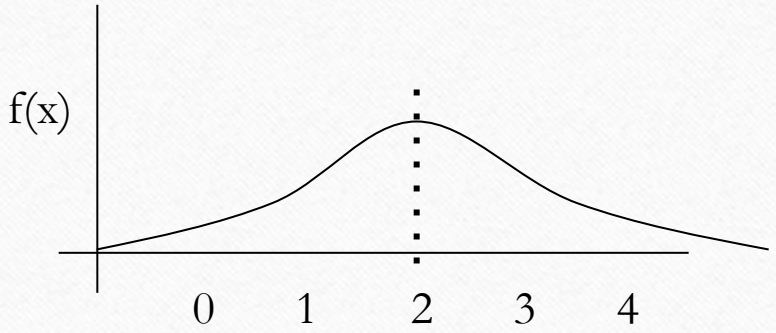
- The concept of randomness or random assignment plays a role in the area of experimental design.

Sampling Terms

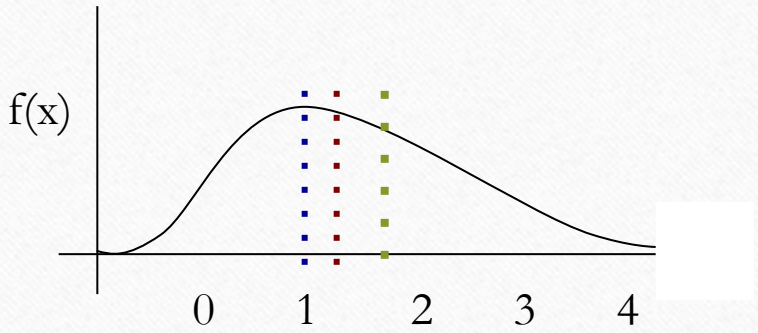
Samples:	Collections of observations
Populations:	Collections of ALL individuals or items of a particular type
Variation:	Change from one observation to another
Variability:	Measure of degree of variation about the mean
Descriptive statistics:	Set of single number statistics that describe a population, such as average, median, standard deviation

Symmetrical Vs Skewed Data

- Symmetrical



- Skewed
- Mean, mode, and median



Skewness of Data

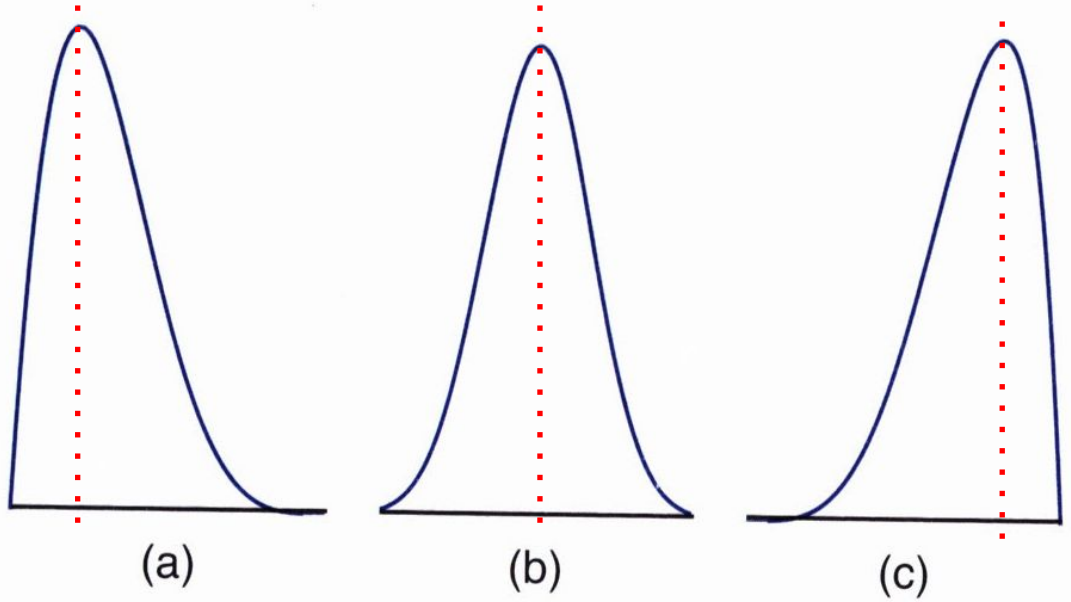


Figure 1.8: Skewness of data.

Skewness?

Measures of Location: Sample Mean

- Suppose that the observations in a sample are $x_1, x_2, \ldots, x_n$
- The sample mean, denoted by $\bar{x}$

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Measures of Location: Sample Median

- The purpose of the sample median is to reflect the central tendency of the sample in such a way that it is uninfluenced by extreme values or outliers.
- Suppose that the observations in a sample are $x_1, x_2, \ldots, x_n$
- The sample median, denoted by $\tilde{x}$

$$\tilde{x} = \begin{cases} x_{n+1/2}, & if\ n\ is\ odd \\ \frac{1}{2}\left(x_{n/2} + x_{n/2+1}\right), & if n\ is\ even \end{cases}$$

Measures of Location: Trimmed Means

- A trimmed mean is computed by "trimming away" a certain percent of both the largest and smallest set of values.
- E.g., the 10% trimmed mean is found by eliminating the largest 10% and smallest 10% and computing the average of the remaining values.
- The trimmed means, denoted by $\bar{x}_{tr(\text{trimmed percent})}$

$$10\%\ trimmed\ means \rightarrow \bar{x}_{tr(10)}$$

Sample Range

$$Sample\ range = X_{max} - X_{min}$$

Q: What is the sample range for the following data?

18.71	21.41	20.72	21.81	19.29	22.43	20.17
23.71	19.44	20.50	18.92	20.33	23.00	22.85
19.25	21.77	22.11	19.77	18.04	21.12.	

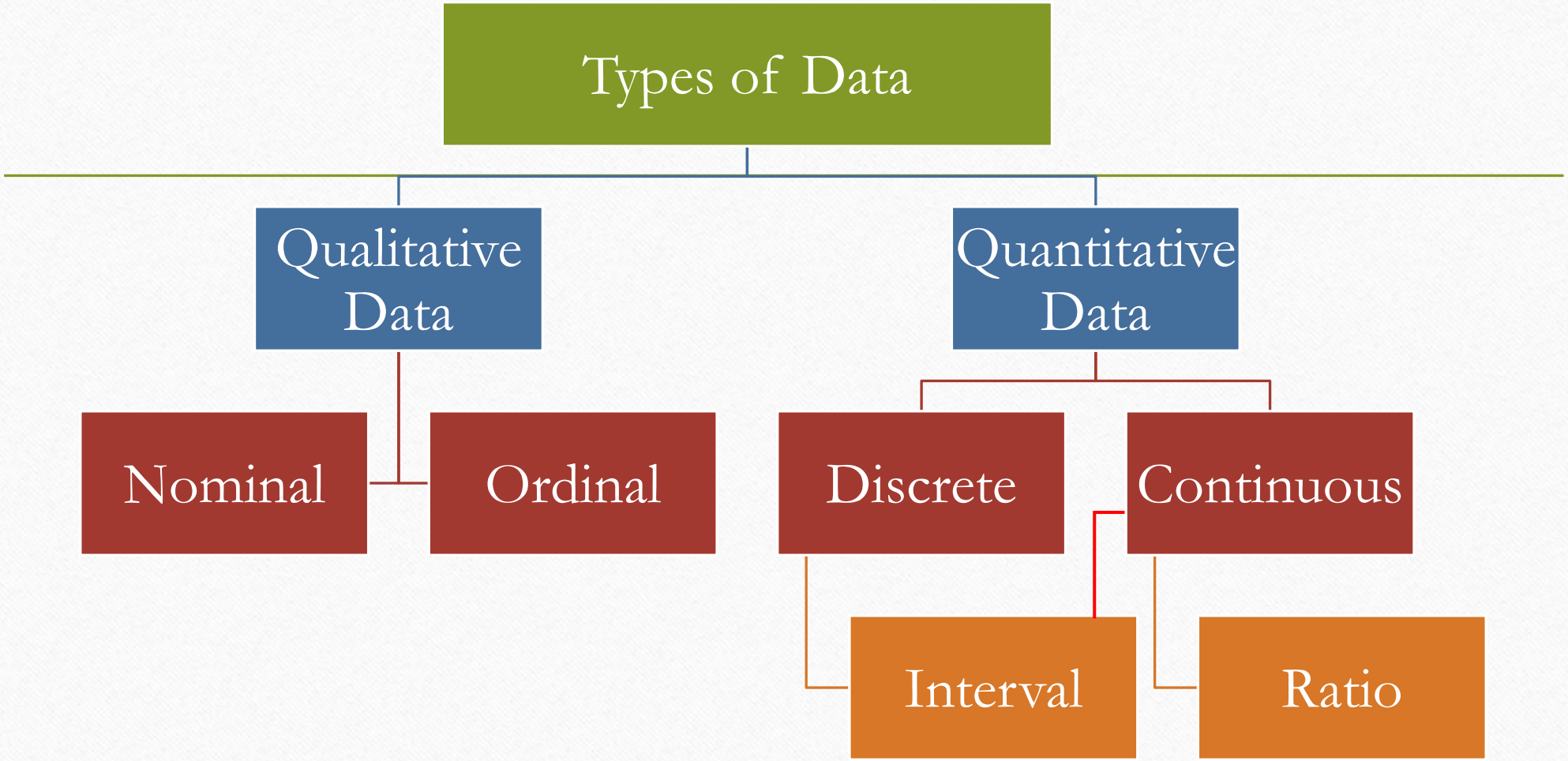
Sample Standard Deviation

- Suppose that the observations in a sample are $x_1, x_2, \ldots, x_n$
- The sample variance, denoted by s^2

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

- The sample standard deviation, denoted by s

$$s = \sqrt{Sample\ Variance}$$


Types of Data
Qualitative Data
Quantitative Data
Nominal
Ordinal
Discrete
Continuous
Interval
Ratio

Level of Measurement

Categorical (entities are divided into distinct categories):

- **Binary variable**: There are only two categories.
- **Nominal variable**: There are more than two categories.
- **Ordinal variable**: The same as a nominal variable but the categories have a logical order.

Continuous (entities get a distinct score):

- **Interval variable**: Equal intervals on the variable represent equal differences in the property being measured.
- **Ratio variable**: The same as an interval variable, but the ratios of scores on the scale must also make sense.

Examples of the basic variable types

Quantitative / Numerical		Qualitative / Categorical	
Continuous	**Discrete**	**Nominal**	**Ordinal**
e.g. Hair length	e.g. DLQI score	e.g. Hair color	e.g. Hair loss
PASI score	Number of failures	Skin texture	Fitzpatrick skin type
Body weight	Age	Acne type	Wagner ulcer grading
VAS pain score		Ethnicity	Weight category
Age			

- Abbreviations: DLQI = Dermatology life quality index; PASI = Psoriasis Area and Severity Index; VAS = Visual Analog Scale

Source: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4763618

The Research Process