Statistical Methods

Descriptive Statistics

Shehu, Nathan Y.

Research Methodology Training by The African Centre for Phytomedicine Research and Development, University of Jos. 24th-28th June, 2019

Learning objectives

- **To have basic understanding of concepts used in statistical methods**
- **To make a good sense/interpretation of frequently used statistical tests**
- **To have hands on experience of data analysis**

Statistical methods

- **Mathematical**
  - **formulae**
  - **models and**
  - **techniques**
- **Used in statistical analysis of raw research data.**

Statistics

- **Branch of science that deals with:**
  - **collection**
  - **Organization**
  - **analysis of data**
  - **drawing of inferences from the samples to the whole population.**

Data

- **Recorded facts or pieces of information that can be analyzed or used in an effort to gain knowledge or make decisions.**

Steps in Research Process

Data

- **Nominal – categories that do not have a natural order, e.g. gender, eye colour, types of building**
- **Ordinal – categories which have a natural order but are not numerical, e.g. Likert scales, severity(mild moderate, severe)**
- **continuous – numerical data ordered against a constant scale, e.g. date, temperature, length, weight, frequency**

What is Descriptive Statistics

- **Summarizing and organizing the data so they can be easily understood.**
- **Describes the data, but do not attempt to make inferences from the sample to the whole population. Here, we typically describe the data in a sample.**
- **Is not developed on the basis of probability theory.**

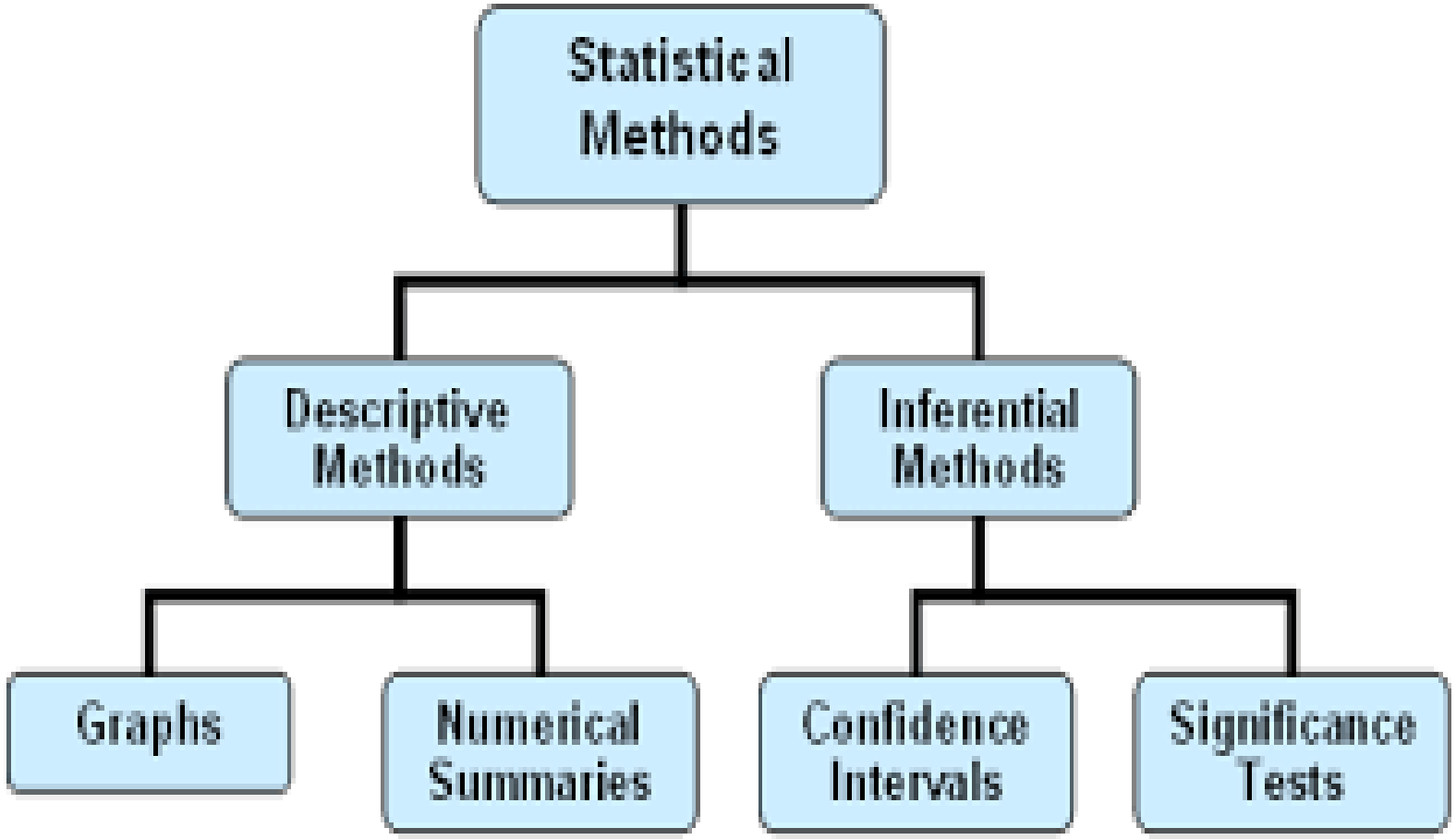
Inferential Statistics

- **Using data collected from samples of a population, and making generalizations which go beyond the sample to the population from which they came.**

Descriptive Statistics types

- **Measures of central tendency**

- **Measures of variability (spread).**


Statistical Methods
Descriptive Methods
Inferential Methods
Graphs
Numerical Summaries
Confidence Intervals
Significance Tests

Types of descriptive statistics:

- **Measures of Central Tendency: Mean, Median, and Mode. ...**
- **Measures of Frequency: Count, Percent, Frequency. ...**
- **Measures of Dispersion or Variation: Range, Variance, Standard Deviation. ...**
- **Measures of Position: Percentile Ranks, Quartile Ranks.**

Measures of central tendency: Mean

- **A number around which a whole data is spread out.**
- **A single number which can estimate the value of whole data set.**
- **Calculate mean age(yrs) of the data set having 8 integers: 10, 23, 54, 43, 71, 10, 11, 42**
- **Mean = $\frac{10 + 23 + 54 + 43 + 71 + 10 + 11 + 42}{8}$**
- **Mean = 33 years**

Measures of central tendency: Midian

- **Is the value which divides the data in 2 equal parts when data is arranged in either ascending or descending order.**
- **Median is the middle term, if number of items is odd**
  - **A. What is the median age (yrs) of pupils in a sample from a primary school is 8, 10, 11, 12, 13, 14, 16.?**
    - **8, 10, 11, 12, 14, 15, 16**
    - **8, 10, 11, 12, 14, 15, 16, 18?**

Measures of central tendency: Mode

- **Item with highest frequency of occurrence in a data set.**
- **Exams score: 25, 35, 55, 67, 67, 70, 80. Mode?**
    - **67**
- **Nil mode: equal frequency**
- **Bimodal: 2 values appeared same time and more than the rest of the values**
- **Trimodal: 3 values appeared same time and more than the rest of the values**
- **Multimodal: n modes.**

Measures of dispersion (spread): Standard deviation (SD)

- **Measurement of average distance between each item and the mean.**
- **Sample Standard Deviation:**

$$\text{S.D.} = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n}(x-\bar{x})^2}$$

- **Population (SD):**
  - **μ = population mean**

$$\text{S.D.} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(x-\mu)^2}$$

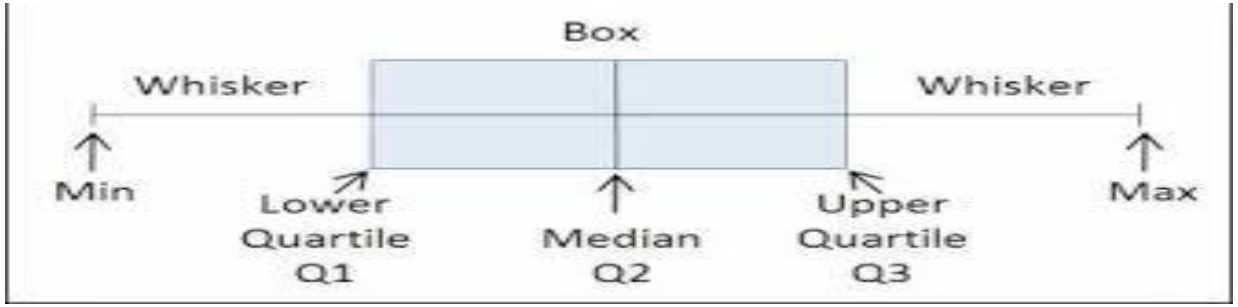
Measures of dispersion (spread): Variance

- **Is the square of average distance between each quantity and mean.**
- **It is the square of standard deviation.**

Measures of dispersion (spread): Range

- **Range: It is the difference between lowest and highest value. 8, 10, 11, 12, 13, 14, 15, 16, 18, 19**
  - **e.g 8-19?**
- **Percentile : nth percentile is n/100 X n**
  - **30^{th} percentile is 30/100 X 10 >>> 3^{rd} position=11**
- **if k is nth percentile, n% of the total terms are less than k.**
  - **e.g 11 is 30^{th} percentile, 30% of the total terms are less than 13**

Measures of dispersion (spread): Quartiles

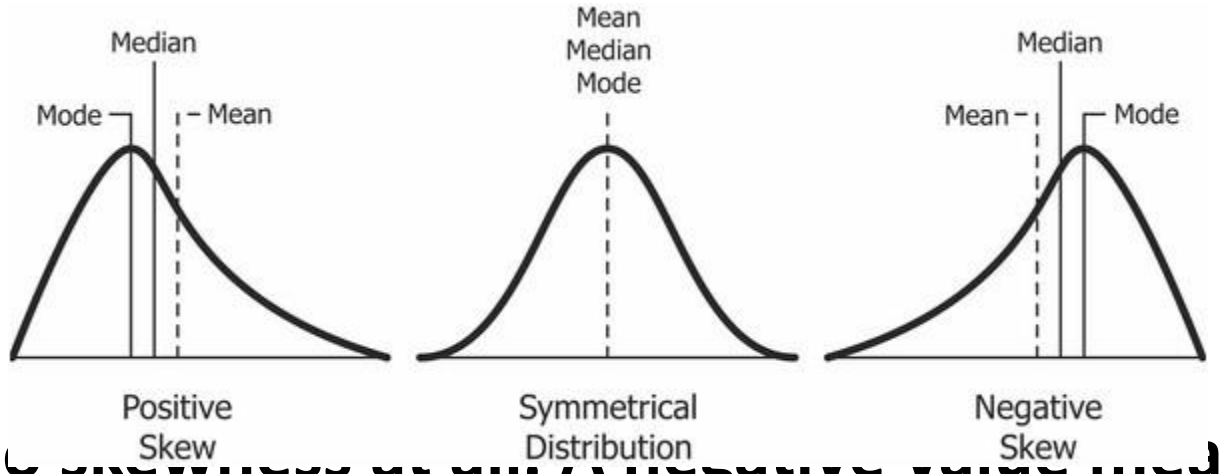


- **data is sorted in an ascending order.**
- **First quartile value is at 25 percentile, second quartile is 50 percentile and third quartile is 75 percentile.**
- **Interquartile range (IQR) = Q3 - Q1**

Measures of dispersion (spread): Skewness

- **In a perfect normal distribution, the tails on either side of the curve are exact mirror images of each other.**
- **Negative skewness: the tail on the curve's left-hand side is longer than the tail on the right-hand side, and the mean is less than the mode.**
- **positive skewness: the tail on the curve's right-hand side is longer than the tail on the left-hand side, and the mean is greater than the mode. This situation is also called.**
-

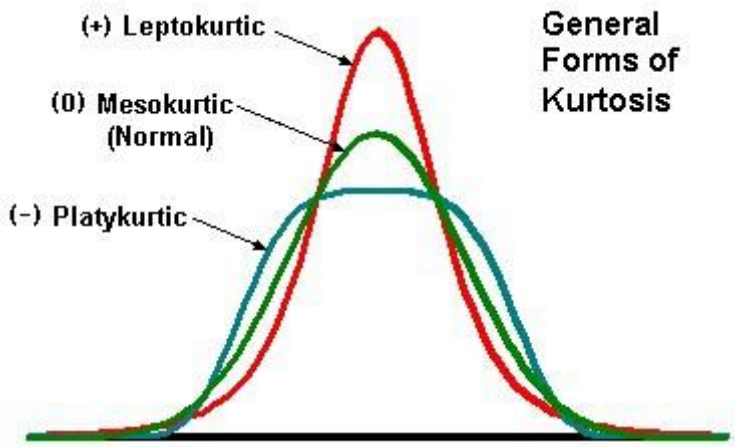
Measures of dispersion (spread): Skewness

- 

- **Zero means no skewness at all. A negative value means the distribution is negatively skewed. A positive value means the distribution is positively skewed**

Measures of dispersion (spread): Kurtosis

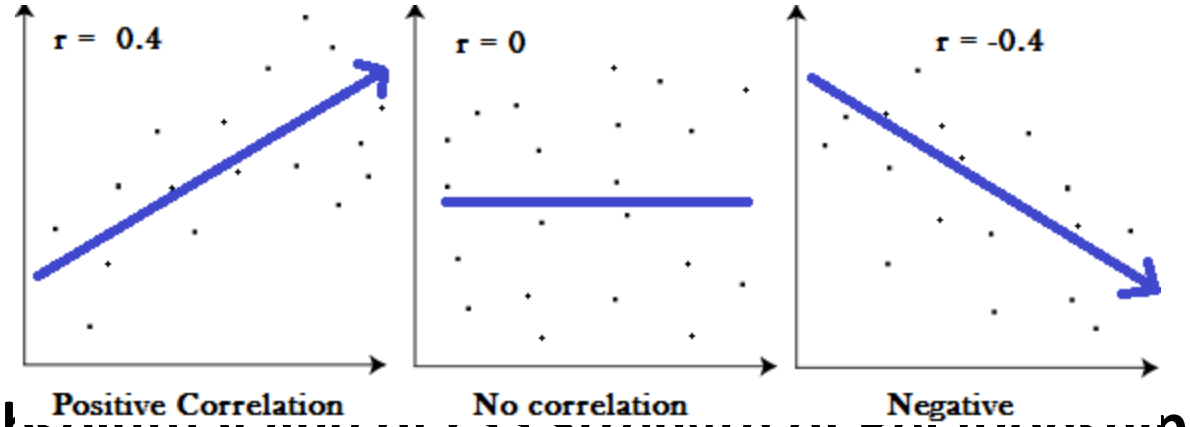
- **The existence of outliers.**
- **measure of whether the data are long-tailed or short-tailed relative to a normal distribution.**



Correlation

- **Technique that shows how strongly pairs of variables are related.**
- **Correlation coefficient (“r”). It ranges from -1.0 to +1.0.**
  - **The closer r is to +1 or -1, the more closely the two variables are related.**
  - **If r is positive, it means that as one variable gets larger the other gets larger.**
  - **If r is negative it means that as one gets larger, the other gets smaller (“inverse” correlation).**
  - **If r is close to 0, it means there is no relationship between the variables.**

Correlation



- Absolute value of r>> Strength of Relationship
- **r < 0.3: None or very weak**
- **0.3 < r <0.5: Weak**
- **0.5 < r < 0.7: Moderate**
- **r > 0.7: Strong**

Comments and Questions?