Biostatistics

Lecture 4

Prepared by Dr. Ibrahim AL-Jaafari

Descriptive Statistics
Measures of Central Tendency

Mean

*** For Ungrouped Data**

Mean = The sum of all values ÷ Number of values

Note : The mean is used to represent the average when the data are normally distributed or symmetrical shaped.

Descriptive Statistics
Measures of Central Tendency

Mean

*** For Ungrouped Data**

Example / Find the mean of the following data (158, 189, 265, 127, 191)

$$158 + 189 + 265 + 127 + 191 = 930$$

$$930 \div 5 = 186$$

Descriptive Statistics
Measures of Central Tendency

Mean

*** For Grouped Data**

$$\text{Mean} = \Sigma mf/n$$

Σ = Sum

F = Frequency

M = Midpoint of the class

n = Total of frequencies

Descriptive Statistics
Measures of Central Tendency

Mean

*** For Grouped Data**

Example / Find the mean of the following grouped data

Age	Frequency
10 - 12	4
13 - 15	12
16 - 18	20
19 - 21	14
	Total = 50

Descriptive Statistics
Measures of Central Tendency

Mean

*** For Grouped Data**

Solution / Mean = $\Sigma mf/n = 832 / 50 = 16.64$

Age	Frequency	Class Midpoint	Midpoint × Frequency
10 - 12	4	11	11× 4 = 44
13 - 15	12	14	14× 12 = 168
16 - 18	20	17	17 × 20 = 340
19 - 21	14	20	20 × 14 = 280
	Total = 50		Total = 832

Descriptive Statistics
Measures of Central Tendency

Mean

Advantages and disadvantages of using the mean
The most common measure of central tendency
Easy to use
It is effected by adding or deleting a new values
It is effected by outlier or exterem values
No need to order the data
It takes into account the whole values to be calculated
The sum of deviations from the mean must be equal to zero

Descriptive Statistics
Measures of Central Tendency

Median

Median = The midpoint value of a set of ordered data from lowest to highest

How to calculate median ?

1-) If the number of values is odd.

Order the values from lowest to highest

Find the location of the median $Median\ Location = \frac{N+1}{2}$

Descriptive Statistics
Measures of Central Tendency

Median

Example:

Find the median for the following values (1, 6, 5, 2, 4)

Order the data (1, 2, 4, 5, 6)

The location of the median = 5 + 1 / 2 = 3 (The third value of the data set)

Median = 4

Descriptive Statistics
Measures of Central Tendency

Median

2-) If the number of values is even.

Order the values from lowest to highest

Find the location of the median = Number of values + 1 / 2

Descriptive Statistics
Measures of Central Tendency

Median

Example:

Find the median for the following values (7, 1, 6, 5, 2, 4)

Order the data (1, 2, 4, 5, 6, 7)

The location of the median = 6 + 1 / 2 = 3.5

Median = 4 + 5 / 2 = 4.5

Descriptive Statistics
Measures of Central Tendency

Median

Advantages and disadvantages of using the median
Easy to use
It is not effected by adding or deleting a new values
It is not affected by outlier or extreme values
The data should be ordered from lowest to the highest
It does not take into account the whole values to be calculated
The sum of deviations from the mean should not be equal to zero

Note : The median is used to represent the average when the data are not normally distributed or symmetrical, for instance the skewed distribution to the right or to the left.

Descriptive Statistics
Measures of Central Tendency

Mode

Mode is the most frequent value in the distribution, there is no way to calculate the mode.

Uni-modal can be found when the data have only one peak, just one mode in the distribution.

Bi-modal can be found when the data have two peaks, this means two modes in the distribution.

Multi-modal can be found when the data have more than two peaks, it means more than two modes.

Descriptive Statistics
Measures of Central Tendency

Mode

Example:

Find the mode for the following data ?

77 ،69 ، 74 ،81 ،71 ،68 ، 74 ،73

Mode = 74

Descriptive Statistics
Measures of Central Tendency

Mode

Advantages and disadvantages of using the mode
Easy to use
It is not effected by adding or deleting a new values
It is not effected by outlier or exterem values
The data should not be ordered from lowest to the highest
It does not take into account the whole values to be calculated
The sum of deviations from the mean should not be equal to zero

Descriptive Statistics
Measures of Central Tendency

The relationship between mean, median and mode

1-) The data is normally distributed (Symmetry)

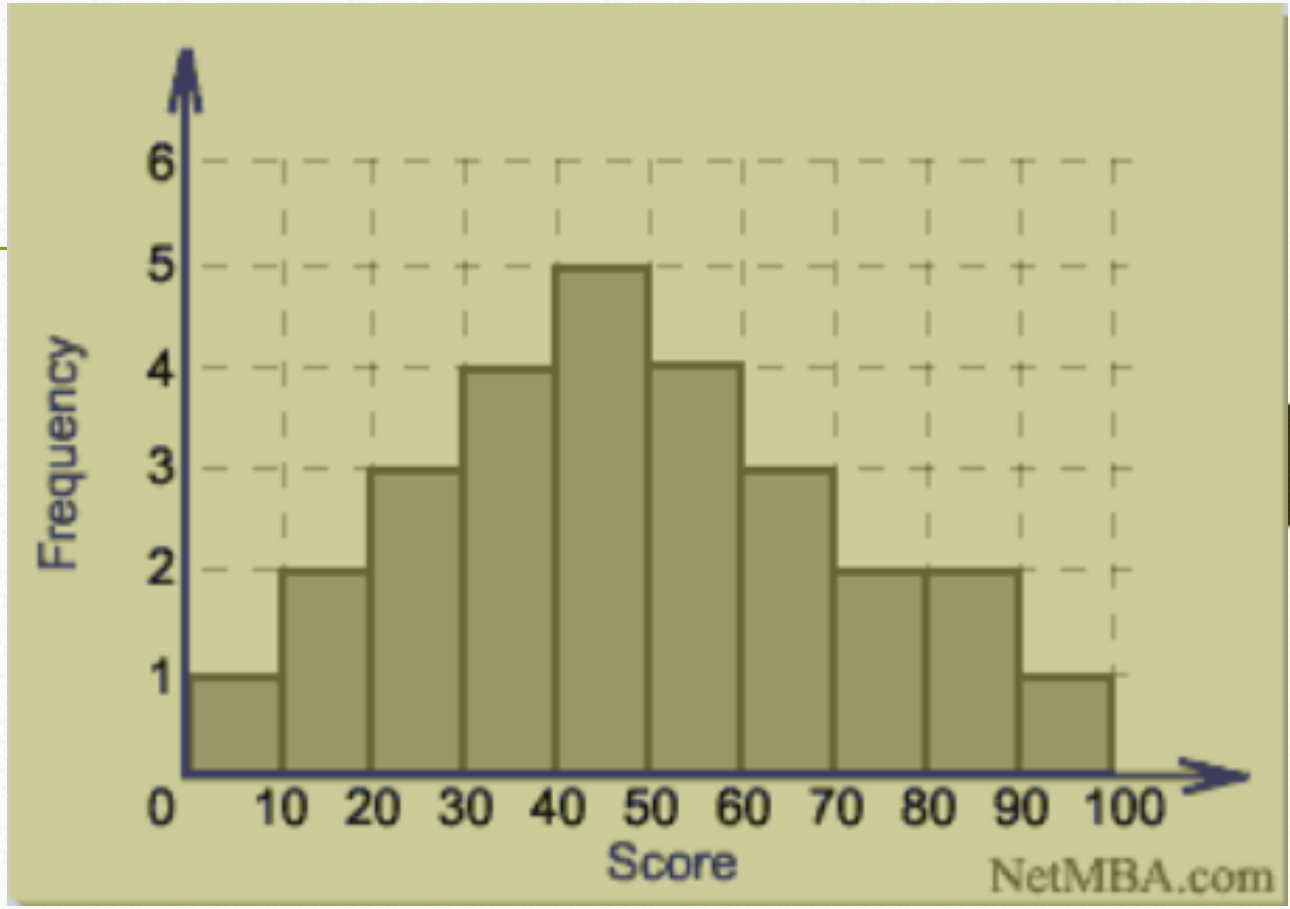
The values of mean, median and mode are very close to each other in

Uni-modal and the shape of the distribution is symmetrical (bell shaped).

Descriptive Statistics Measures of Central Tendency

The relationship between mean, median and mode

1-) The data is normally distributed (Symmetry)



Descriptive Statistics
Measures of Central Tendency

The relationship between mean, median and mode

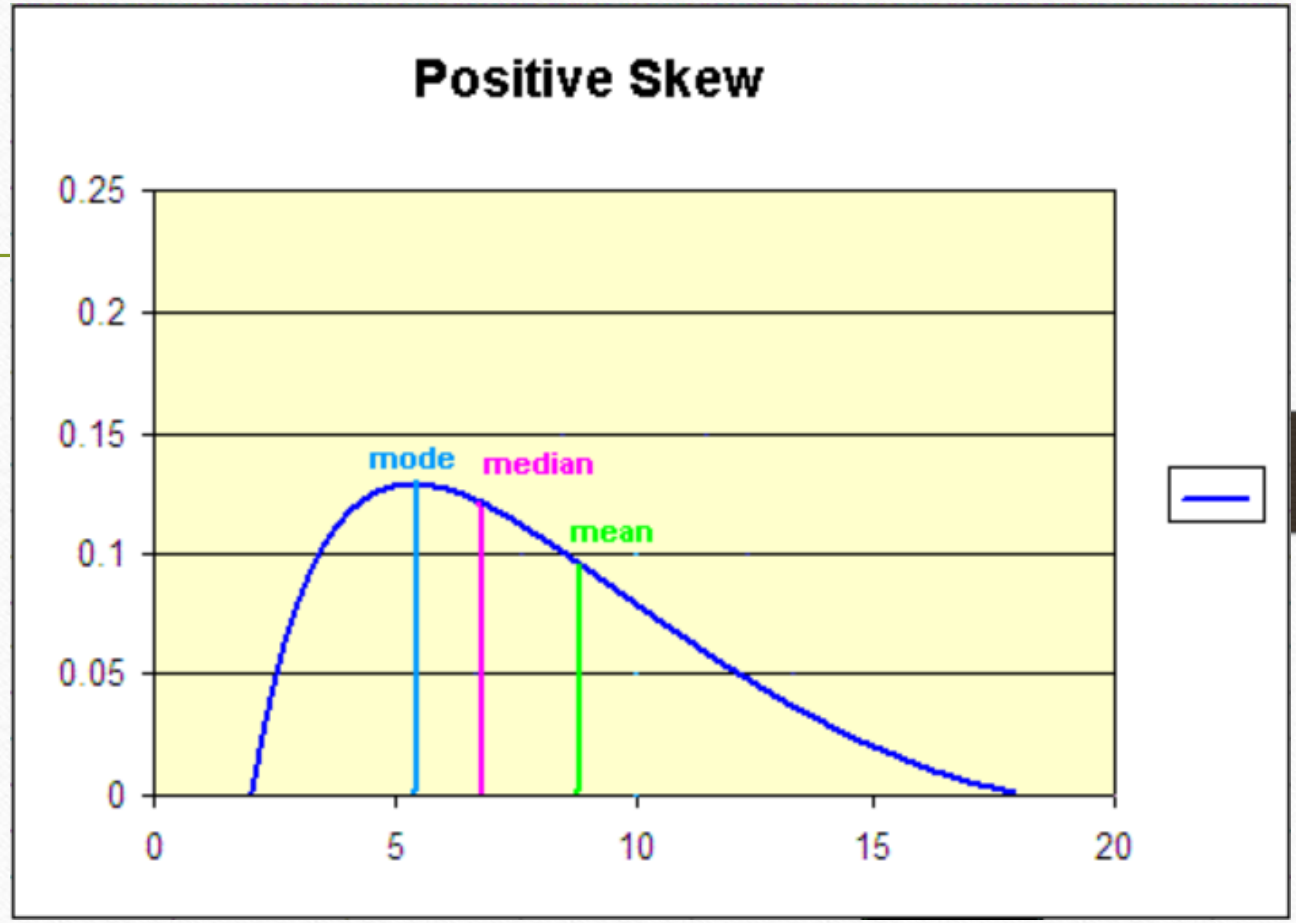
2-) The data is skewed to the right (positive skewed)

The mean is greater than median and the tail is longer on the right side.

Descriptive Statistics Measures of Central Tendency

The relationship between mean, median and mode

2-) The data is skewed to the right (positive skewed)



Descriptive Statistics
Measures of Central Tendency

The relationship between mean, median and mode

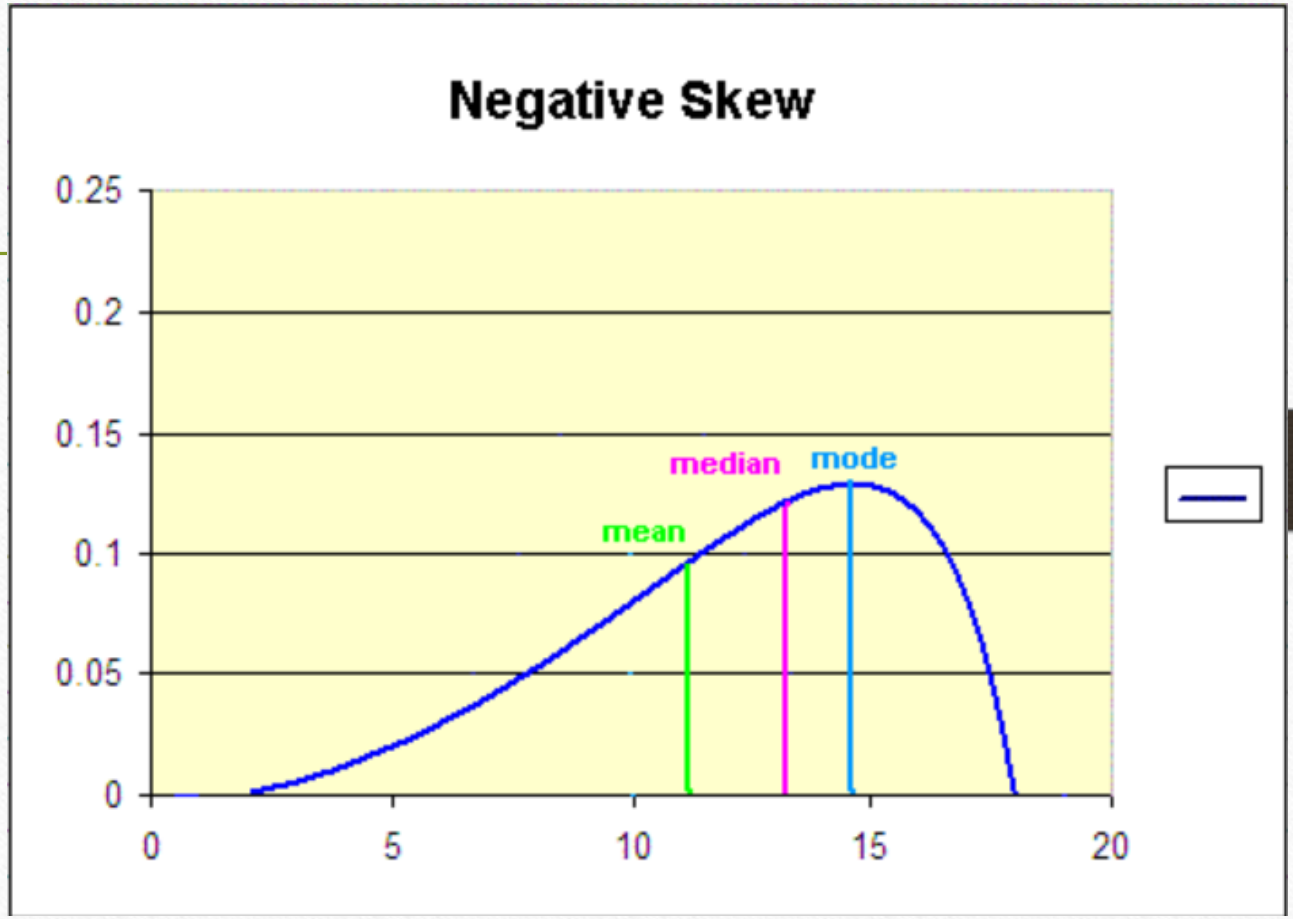
3-) The data is skewed to the left (negative skewed)

The mean is lower than median and the tail is longer on the left side.

Descriptive Statistics Measures of Central Tendency

The relationship between mean, median and mode

3-) The data is skewed to the left (negative skewed)



Reference

- **Prem S. Mann 1998, Introductory Statistics, 7th edn, New York, USA.**

Descriptive Statistics
Measures of Dispersion

The measures of central tendency, such as the mean, median, and mode, do not reveal the whole picture of the distribution of a data set.

Two data sets with the same mean may have completely different spreads.

Descriptive Statistics
Measures of Dispersion

- Mean for Group 1 = 728 / 10 = 72.8

- Mean for Group 2 = 728 / 10 = 72.8

100	90	88	85	80	75	70	55	45	40	Group 1
78	77	76	75	74	73	72	70	68	65	Group 2

Descriptive Statistics
Measures of Dispersion
(Range)

- Range for Group 1 = 100 – 40 = 60

  Group 1 has large variation

- Range for Group 2 = 78 – 65 = 13

  Group 2 has small variation

Range = Largest value - Smallest value

100	90	88	85	80	75	70	55	45	40	Group 1
78	77	76	75	74	73	72	70	68	65	Group 2

Descriptive Statistics
Measures of Dispersion
(Variance & Standard Deviation)

- The standard deviation is the most-used measure of dispersion. The value of the standard deviation tells how closely the values of a data set are spread around the mean.
- Lower value of the standard deviation = Smaller range around the mean.
- Larger value of the standard deviation = Larger range around the mean.

Descriptive Statistics
Measures of Dispersion
(Variance & Standard Deviation)

- How to calculate variance and standard deviation ??

- Variance = $$s^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$$

- Standard Deviation = $$s = \sqrt{\frac{\sum(x - \overline{x})^2}{N-1}}$$

Descriptive Statistics
Measures of Dispersion
(Variance & Standard Deviation)

- Example / Calculate variance and standard deviation for the following data
- 1, 2, 3, 4, 5
- Answer /
- Mean = 1 + 2 + 3 + 4 + 5 = 15 / 5 = 3

Descriptive Statistics Measures of Dispersion (Variance & Standard Deviation)

X	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
1	1 – 3 = -2	4
2	2 – 3 = -1	1
3	3 – 3 = 0	0
4	4 – 3 = 1	1
5	5 – 3 = 2	4
Total	**= 0**	**10**

Variance = 10 / 5 – 1

Variance = 10 / 4 = 2.5

Standard deviation = Square root of variance

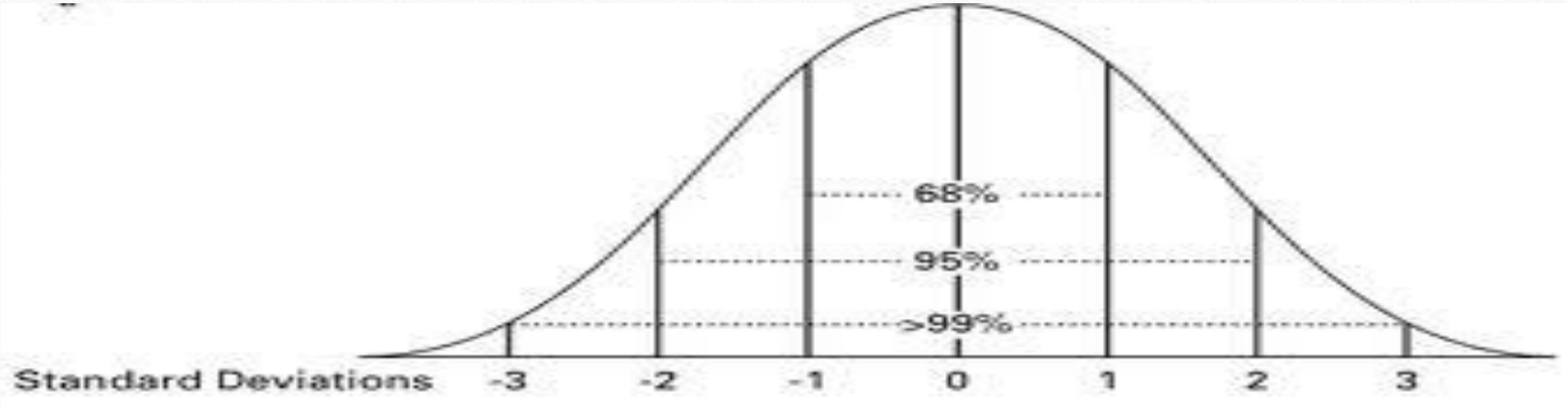
$\sqrt{}$ 2.5 = 1.58

- Mean = 3

Descriptive Statistics
Measures of Dispersion
(Variance & Standard Deviation)

Advantages and disadvantages of using standard deviation
The most common measure of dispersion
Easy to use
It is effected by outlier or exterem values
It takes into account the whole values to be calculated

Descriptive Statistics
Measures of Dispersion
(Variance & Standard Deviation)



Descriptive Statistics Measures of Dispersion (Variance & Standard Deviation)

- If we have a group of students had a normal distribution for weight. The mean weight was 70kg and the SD was 5kg.

1 SD above the mean weight is 70 + 5 = 75kg

1 SD below the mean weight is 70 - 5 = 65kg

2 SD above the mean weight is 70 + 10 = 80kg

2 SD below the mean weight is 70 - 10 = 60kg

3 SD above the mean weight is 70 + 15 = 85kg

3 SD below the mean weight is 70 - 15 = 55kg

1 SD will include 68.2% of students weigh between (75 and 65 kg)

2 SD will include 95.4% of students weigh between (80 and 60 kg)

3 SD will include 99.7% of students weigh between (85 and 55 kg)

Reference

- **Prem S. Mann 1998, Introductory Statistics, 7th edn, New York, USA.**

Inferential Statistics
Hypothesis Test About the Mean

- There are two types of hypothesis:-

1- Null Hypothesis (H0) / It is a claim (or statement) about a population parameter that is assumed to be true until it is declared false.

Null hypothesis is denoted by : H0

Inferential Statistics
Hypothesis Test About the Mean

- **There are two types of hypothesis:-**

1- Null Hypothesis (H0) /**It means there is no difference or relationship between the variables. If there is a significant difference or relationship between variables, we would reject the null hypothesis and accept alternative hypothesis. while, If there is no significant difference or relationship between variables, then we would accept the null hypothesis and reject alternative hypothesis.**

Inferential Statistics
Hypothesis Test About the Mean

- There are two types of hypothesis:-

1- Alternative Hypothesis (H1) or (Ha) / It is a claim about a population parameter that will be true if the null hypothesis is false.

Alternative Hypothesis is denoted by : H1 or Ha

Inferential Statistics
Hypothesis Test About the Mean

- There are two types of hypothesis:-

1- Alternative Hypothesis (H1) or (Ha) /

It means a research hypothesis. This is what the researcher want to prove in his research.

Inferential Statistics
Hypothesis Test About the Mean

- There are two types of errors in hypothesis test:-

1- Type I Error or (α Error):- It means reject null hypothesis when it is true, the probability of making type I error is call Alpha, it can be determined by level of significance **(0.01, 0.05 and 0.10) = (1%, 5% and 10%)**

Inferential Statistics
Hypothesis Test About the Mean

- There are two types of errors in hypothesis test:-

1- Type II Error or (ß Error):- It means accept null hypothesis when it is false, the probability of making type II error is call Beta, it is related to power of the test (1 - Beta) which means the probability of detecting the real difference or relationship between variables so, the power of the test would be increased if we accept the true null hypothesis.

Inferential Statistics
Hypothesis Test About the Mean

- There are two types of errors in hypothesis test:-

Decision	True	False
Accept	Correct	Type II error
Reject	Type I error	Correct

Inferential Statistics
Hypothesis Test About the Mean

- **Example of types of errors in hypothesis test:-**

		Actual Situation	
		The Person Is Not Guilty	The Person Is Guilty
Court's decision	The person is not guilty	Correct decision	Type II or β error
	The person is guilty	Type I or α error	Correct decision

Inferential Statistics
Hypothesis Test About the Mean

- If the P-value is less than < significance level (0.01 or 0.05 or 0.10) then reject null hypothesis and accept alternative hypothesis.

There are enough evidences to reject H0 and accept H1

- If the P-value is greater than > significance level (0.01 or 0.05 or 0.10) then accept null hypothesis and reject alternative hypothesis

There are less evidences to reject H0 and accept H1

Reference

- **Prem S. Mann 1998, Introductory Statistics, 7th edn, New York, USA.**

Good Luck for All Students

- Please do not hesitate to contact me if you have any questions.
- Dr. Ibrahim AL-Jaafari
- www.Alghamdi-Biostatistics.com
- Email. Bio-stat@Hotmail.com
- Mobile Number : 0553777925