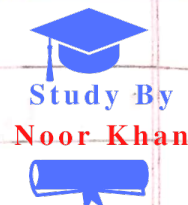


STA301 Mid Term Preparation



Sampling error = $\bar{X} - \mu$

Population mean:

$$\mu = \frac{\sum fx}{\sum f}$$

Range:

$$\text{Range} = X_{\max} - X_{\min}$$

Class interval:

$$h = \frac{\text{Range}}{\text{No. of classes}}$$

No. of classes

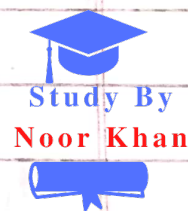
⇒ In case of Raw Data Pertaining to a Continuous Variable

Mode: Mode is the value which occurs most frequently in a set of data.

Example:

Data : 2, 7, 9, 5, 8, 9, 10, 9

Mode of data : 9



2.
⇒ The Mode in case of the Frequency Distribution of a Continuous Variable

Mode:

⇒

$$\hat{X} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where,

l = Lower class boundary of the modal class,

f_m = Frequency of the modal class,

f_1 = Frequency of the class preceding the modal class.

f_2 = Frequency of the class following modal class

h = Length of class interval of the modal class.

Arithmetic Mean:

Def: "The arithmetic mean or simply the mean is a value obtained by dividing the sum of all the observations by their number."

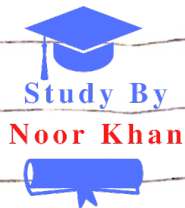
Formula:

$$\bar{X} = \frac{\text{Sum of all the observations}}{\text{Number of the observations}}$$

$$\bar{X} = \frac{\sum X}{n}$$

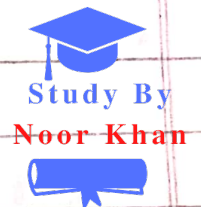
⇒ In case of frequency distribution

Arithmetic Mean:



$$\bar{X} = \frac{\sum f \cdot X}{\sum f} = \frac{\sum fX}{n}$$

Weighted Mean:



$$\bar{X}_w = \frac{\sum WX}{\sum W}$$

⇒ In case of a frequency distribution of a continuous

Median:

Formula:

$$\tilde{X} = l + \frac{h}{f} \left(\frac{n}{2} - c \right)$$

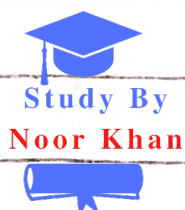
Where,

l = lower class boundary of the median class

h = Class interval size of the median class.

f = Frequency of the median class

$n = \sum f$ (the total number of observations)



c - Cumulative frequency of the class
preceding the median class

Empirical Relation BTW the Mean, Median and the Mode

$$\text{Mode} = 3\text{Median} - 2\text{Mean}$$



Study By

Noor Khan



Quartiles: The quartiles, together with the median, achieve the division of the total area into four equal parts.

1. First Quartile:

$$Q_1 = l + \frac{h}{f} \left(\frac{n}{4} - c \right)$$



Study By

Noor Khan



2. Second Quartile (Median)

$$Q_2 = l + \frac{h}{f} \left(\frac{2n}{4} - c \right)$$

$$\therefore Q_2 = \bar{X}$$

$$Q_2 = l + \frac{h}{f} \left(\frac{n}{2} - c \right)$$

3. Third Quartile

$$Q_3 = l + \frac{h}{f} \left(\frac{3n}{4} - c \right)$$



Study By

Noor Khan

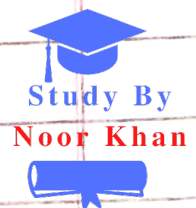


Deciles & Percentiles

The deciles and the percentiles give the division of the total area into 10 and 100 equal parts respectively.

First decile:

$$D_1 = l + \frac{h}{f} \left(\frac{n}{10} - c \right)$$



Second decile:

$$D_2 = l + \frac{h}{f} \left(\frac{2n}{10} - c \right)$$

Third decile:

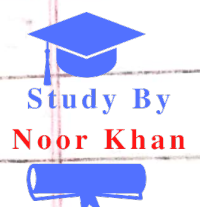
$$D_3 = l + \frac{h}{f} \left(\frac{3n}{10} - c \right)$$

and so on.

$$\therefore D_5 = \bar{X} \text{ (Median)}$$

First percentile

$$P_1 = l + \frac{h}{f} \left(\frac{n}{100} - c \right)$$

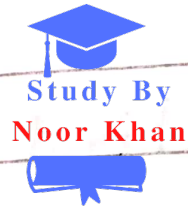


Second percentile:

$$P_2 = l + \frac{h}{f} \left(\frac{2n}{100} - c \right)$$

Third percentile:

$$P_3 = l + \frac{h}{f} \left(\frac{3n}{100} - c \right)$$



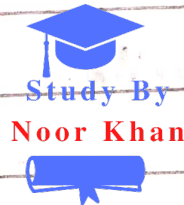
and so on.

MCQ's

$$P_{50} = \tilde{X}$$

$$P_{25} = Q_1$$

$$P_{75} = Q_3$$



$$P_{100} = D_u$$

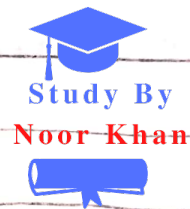
Geometric mean:

The geometric mean, G_1 , of a set of n positive values $X_1, X_2, \dots, X_n$ is defined as the positive n th root of their product.

Formula:

$$G_1 = \sqrt[n]{X_1 X_2 \dots X_n}$$

When n is large



Formula:

$$G = \text{antilog} \left[\frac{\sum \log X}{n} \right]$$

For Grouped data

Formula:

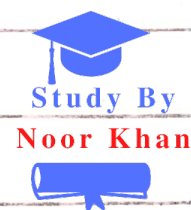
$$G = \text{antilog} \left[\frac{\sum f \log X}{n} \right]$$

Harmonic Mean: The harmonic mean is defined as the reciprocal of the arithmetic mean of the reciprocals of the values.

⇒ In case of raw data:

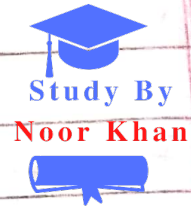
Formula:

$$H.M = \frac{n}{\sum \left(\frac{1}{x} \right)}$$



⇒ In case of grouped data (data grouped into a frequency distribution):

Formula:



$$H.M = \frac{n}{\sum f\left(\frac{1}{x}\right)}$$

(Where x represents the midpoints of the various classes)

V. imp Relation between Arithmetic, Geometric and Harmonic Means

Arithmetic Mean $\geq$ Geometric Mean $\geq$ Harmonic Mean

Five most well-known

Measures of central tendency

Arithmetic mean

Median

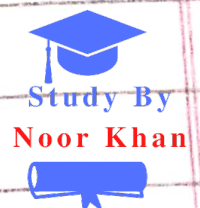
Mode

Geometric Mean

Harmonic Mean

Mid range

Mid Quadile range



Mid-Range:

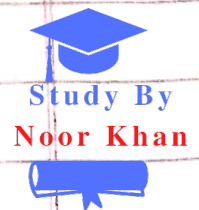
Formula:

$$\text{Mid-range} = \frac{X_o + X_m}{2}$$

Where,

X_o = Smallest observation

X_m = Largest observation



Mid-Quartile Range:

Formula:

$$\text{Mid-quartile range} = \frac{Q_1 + Q_3}{2}$$

Four Measures of Dispersion

Range

The Quartile deviation

The Mean deviation

The Standard deviation

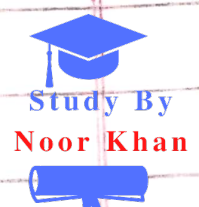
Range:

$$\text{Range} = X_m - X_o$$

where,

X_m = highest value

X_o = lowest value



¹⁰ Coefficient of Dispersion

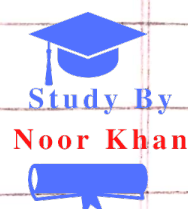
Formula:

$$\text{Coefficient of Dispersion} = \frac{X_m - X_o}{X_m + X_o}$$

Quartile Deviation

Formula:

$$Q.D. = \frac{Q_3 - Q_1}{2}$$



It is also known as Semi-interquartile range.

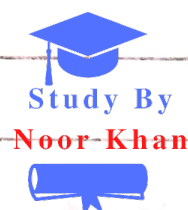
Coefficient of Quartile Deviation

Formula: Coefficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Mean deviation:

(For Ungrouped data)

$$M.D. = \frac{\sum |d|}{n}$$



(For Grouped data)

$$M.D. = \frac{\sum f|d|}{n}$$

$$\because d = x - \bar{x}$$

$$\because \bar{x} = \frac{\sum x}{n}$$

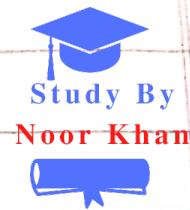
Co-efficient of M.D

$$\text{Co-efficient of M.D} = \frac{\text{M.D}}{\text{Median}} = \frac{\text{M.D}}{\text{Mean}}$$

Variance:

$$\text{Variance} = \frac{\sum (x - \bar{x})^2}{n}$$

Standard Deviation



$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Short cut Formula For Standard Deviation

$$S = \sqrt{\left\{ \frac{\sum x^2}{n} - \left(\frac{\sum x}{n} \right)^2 \right\}}$$

Standard Deviation in Case of Grouped Data

$$S = \sqrt{\frac{\sum f(x - \bar{x})^2}{n}}$$

Short cut Formula of the Standard Deviation in Case of Grouped Data

$$S = \sqrt{\left\{ \frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n} \right)^2 \right\}}$$

Coefficient of S.D

$$\text{Coefficient of S.D} = \frac{\text{Standard Deviation}}{\text{Mean}}$$

Coefficient of Variation

$$C.V = \frac{S}{\bar{X}} \times 100$$

Pearson's Coefficient of Skewness

$$\text{Pearson's Coefficient of Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$$

∴ After applying empirical relation between the mean, median and the mode, the Pearson's Coefficient of Skewness is given by:

$$\text{Pearson's Coefficient of Skewness} = \frac{3(\text{mean} - \text{median})}{\text{Standard deviation}}$$

Percentile Coefficient of Kurtosis

$$K = \frac{Q.D}{P_{90} - P_{10}}$$

Moments

$$m_1 = 0$$

$$m_2 = m'_2 - (m'_1)^2$$

$$m_3 = m'_3 - 3m'_2 m'_1 + 2(m'_1)^3$$

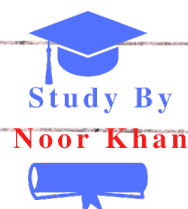
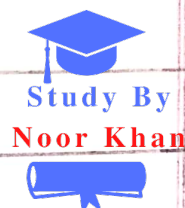
$$m_4 = m'_4 - 4m'_3 m'_1 + 6m'_2 (m'_1)^2 - 3(m'_1)^4$$

Moment Ratios:

$$b_1 = \frac{(m_3)^2}{(m_2)^3}$$

$$b_2 = \frac{m_4}{(m_2)^2}$$

$\therefore$ b_1 is used to measure the skewness of our distribution, and b_2 is used to measure the kurtosis of the distribution



Vimp

11

Equation of a Straight Line

$$Y = a + bX$$

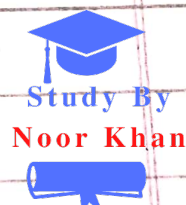
Where,

- Y represents the dependent variable.
- X represents the independent variable.
- a represents the Y -intercept
(i.e. the value of Y when X is equal to zero)
- b represents the slope of the line
(i.e. the value of the $\tan \theta$, where θ represents the angle between the line and the horizontal axis)

imp $\therefore \sum (Y - \hat{Y}) = 0$

Standard Error of Estimate Formula:

$$S_{y.x} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}}$$



Alternative Formula:

$$S_{y.x} = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n-2}}$$

$$S_y = \sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2}$$

Pearson's Coefficient of Correlation

Formula:

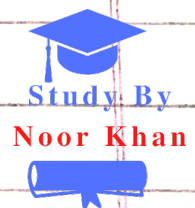
$$r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

Covariance of X and Y

$$\text{Cov}(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n}$$

Short cut Formula:

For the Pearson's Coefficient



$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right) \left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

imp $\Rightarrow$ r is a pure number that lies between -1 and 1

$$-1 < r < 1$$

Permutation

$${}^n P_r = \frac{n!}{(n-r)!}$$

Combination

$${}^n C_r \text{ or } \binom{n}{r}$$

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Binomial expansion

$$(a+b)^n = \sum_{r=0}^n \binom{n}{r} a^{n-r} b^r$$

Probability

$$P(A) = \frac{m}{n} = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}}$$

Law of Complementation

$$P(\bar{A}) = 1 - P(A)$$

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

