BALAJI INSTITUTE OF I.T AND MANAGEMENT KADAPA

STATISTICS FOR MANAGERS

(17E00105)

ICET CODE: BIMK

FIRST INTERNAL

ALSO DOWLOAD AT http://www.bimkadapa.in/materials.html

Name of the Faculty: T.HIMMAT

Units covered: 1st, 2nd & half of 3rd Unit

E-Mail:himmatbimk@gmail.com

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR

MBA I Semester	L	T	P	C
	4	0	0	4

(17E00105) STATISTICS FOR MANAGERS

The objective of this course is to familiarize the students with the statistical techniques popularly used in managerial decision making. It also aims at developing the computational skill of the students relevant for statistical analysis.

1.**Introduction of statistics** – Nature & Significance of Statistics to Business, , Measures of Central Tendency- Arithmetic – Weighted mean – Median, Mode – Geometric mean and Harmonic mean – Measures of Dispersion, range, quartile deviation, mean deviation, standard deviation, coefficient of variation – Application of measures of central tendency and dispersion for business decision making.
2. **Correlation**: Introduction, Significance and types of correlation – Measures of correlation – Co-efficient of correlation. Regression analysis – Meaning and utility of regression analysis – Comparison between correlation and regression – Properties of regression coefficients-Rank Correlation.
3. **Probability** – Meaning and definition of probability – Significance of probability in business application – Theory of probability –Addition and multiplication – Conditional laws of probability – Binominal – Poisson – Uniform – Normal and exponential distributions.
4. **Testing of Hypothesis-** Hypothesis testing: One sample and Two sample tests for means and proportions of large samples (z-test), One sample and Two sample tests for means of small samples (t-test), F-test for two sample standard deviations. ANOVA one and two way .
5. **Non-Parametric Methods:** Chi-square test for single sample standard deviation. Chi-square tests for independence of attributes - Sign test for paired data.

Textbooks:

- Statistical Methods, Gupta S.P., S.Chand. Publications

References:

- Statistics for Management, Richard I Levin, David S.Rubin, Pearson,
- Business Statistics, J.K.Sharma, Vikas house publications house Pvt Ltd
- Complete Business Statistics, Amir D. Aezel, Jayavel, TMH,
- Statistics for Management, P.N.Arora, S.Arora, S.Chand
- Statistics for Management , Lerin, Pearson Company, New Delhi.
- Business Statistics for Contemporary decision making, Black Ken, New age publishers.
- Business Statistics, Gupta S.C & Indra Gupta, Himalaya Publishing House, Mumbai

Subject :
Title of the test case :
Case study No. :
Date :
Page No. :

UNIT - 1

Introduction To Statistics

Meaning of Statistics :-

Statistics means collecting the data, organizing the data, presenting the data, analysing the data and interpreting the data.

Meaning of statistics in 2 Senses. Singular Sense and plural sense.

In Singular Sense :-

In Singular sense, statistics means collection, organizing, presenting, analyzing and interpreting the data.

In Plural Sense :-

* In plural sense, statistics means collection of numerical facts.

* In this sense not only consider figures but also take percentages, averages and co-efficient derived from numerical facts.

Nature and Significance of Statistics to Business :-

Statistics is used, to extended the different fields of experiment to draw valid conclusions, and it is used to found the increase importance and usage. The nature and importance of statistics in various fields are listed given below.

* State Affairs :-
    * To collect the information and study the economic condition of people in the states.
    * To assess the resources available in states.
    * To help state to take decision on accepting (or) rejecting it's policy based on statistics.
    * To provide information and analysis on various factors of state like wealth, crimes, agriculture experts, education etc.,
* Economics :-
    * Helps in formulation of economic laws and policies.
    * Helps in studying economic problems.
    * Helps in compiling the national income accounts.
    * Helps in economic planning.
* Business :-
    * Helps to take decisions on location and size.
    * Helps to study demand and supply.
    * Helps in forecasting and planning.
    * Helps controlling the quality of the products (or) process.
    * Helps in making marketing decisions.
    * Helps for production, planning and inventory management.
    * Helps in business risk analysis.

* Education :-

statistics is necessary to formulate the policies regarding start of new courses, consideration of facilities available for proposed courses.

* Accounts And Audits :-

* Helps to study the correlation between profits and dividends enable to know trend of future profits.
* In auditing sampling techniques are followed.

* Measure of Central Tendency :-

According to Simpson and Kafka, "Measure of central Tendency is a single value with in the range of the entire mass of data that is used to represent the whole data".

Characteristics :-

* It should be strongly defined :-

An average should be strongly defined so that there is no confusion in regard to it's meaning. If it is not well defined it may be influenced by the prejudice value to represent the distribution.

* It's definition should be in the form of a Mathematical formula :-

With mathematical formulation different persons may not interpret it differently and anybody compating the average form a set of data.

* It should be easy to calculate & Simple to follow :-

An average should be simple in comprehension So that it can be calculated with reasonable case and it's use will be very limited.

* It should be based on all observations in the series :-

An average will be truly representative of the whole mass of data if it is computed from all the observations.

* It should be capable of further algebraic treatment :-

An average should extend itself readly to further algebraic treatment.

* It should be capable of being used in further statistical computation of processing :-

An average should possess this quality.

For Eg :- * Arithmetic mean is suitable for calculating standard deviation else, it will not be of great use in further statistical analysis and it's utility will be limited.

Subject :		Date :	
Title of the test case :			
Case study No. :		Page No. :	

* It should possess Sampling stability :-

An average should be least affected by Sampling. By this we mean that if we take independent random samples of the same size from a given population.

Types of Averages :-

The following are the important types of averages.

- * Arithmetic mean.
- * Median.
- * Mode.
- * Geometric Mean.
- * Hormonic Mean.

Arithmetic Mean :-

Simple Arithmetic Mean :-

In Individual Series, the process of computing arithmetic mean is the ratio between sum of the observations to the total number of observations. i.e., Symbolically, it is denoted by

Individual Series by direct Method :-

Arithmetic mean $\boxed{A.M., \bar{x} = \frac{\Sigma x}{N}}$.

Where, $\Sigma x = x_1 + x_2 + x_3 + \text{- - - - - - - - -} + x_n$.

N = Potal Observations.

Individual Series by shortcut Method :-

$$\boxed{A.M., \bar{x} = A + \frac{\Sigma d}{N}}$$

Where, A = Assumed mean.

$d = x - A$

(difference between observation to the assumed mean)

Problems :-

Individual Series :-

1) The monthly income of 5 persons are given below. Find arithmetic mean.

132, 140, 144, 136 and 148.

Sol:- Direct Method :-

Given 5 persons monthly income are 132, 140, 144, 136, 148.

x
132
140
144
136
148
$\Sigma x = 700$

Where, $\Sigma x = 700$

$N = 5$

$$\text{A.M.}, \bar{x} = \frac{\Sigma x}{N} = \frac{132+140+144+136+148}{5}$$

$$= \frac{700}{5}$$

$$= 140.$$

$$\boxed{\therefore \bar{x} = 140.}$$

Short-cut Method :-

Where, Assumed mean A = 140.

x	$d = (x-A)$
132	-8
140	0
144	4
136	-4
148	8
	$\Sigma d = 0$

$$A.M., \bar{x} = A + \frac{\Sigma d}{N}$$

$$= 140 + \frac{0}{5}$$

$$= 140 + 0$$

$$= 140.$$

$$\boxed{\therefore \bar{x} = 140.}$$

Arithmetic in Discrete Series :-

Direct Method :-

(i) Multiply each item/variable to it's frequency i.e., f×x.

(ii) Add all the fx values i.e., Σfx.

(iii) Add sum of all the frequencies i.e., $N = \Sigma f$.

Symbolically, it is denoted by

$$\boxed{A.M., \bar{x} = \frac{\Sigma fx}{\Sigma f.}}$$

Shortcut Method :-

(i) Assume any one of the variable in the given series i.e., A for easy calculations.

(ii) Find the deviation value i.e., $d = x - A$ (difference between variable to the assumed mean)

(iii) Multiply the frequencies with all the deviation values

Subject : Date :

Title of the test case :

Case study No. : Page No. :

i.e., fd. Then add all the values i.e., Σfd.

Sum all the frequencies i.e., Σf.

Symbolically, it is denoted by.

$$\boxed{A.M., \bar{x} = A + \frac{\Sigma fd}{N}}$$

Where, A = Assumed mean.

$d = x - A$.

$N = \Sigma f$ (sum of all frequencies).

Problem:-

1) calculate Arithmetic mean from the following data.

Marks (x)	:	20	30	40	50	60	70
No. of students (f)	:	8	12	20	10	6	4

Sol:- Direct Method:-

Marks (x)	No. of Students (f)	fx
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	N=60	$\Sigma fx = 2460$

$$A.M. \bar{x} = \frac{\Sigma fx}{\Sigma f \text{ or } N}$$

$$= \frac{2460}{60}$$

$$\bar{x} = 41.$$

$$\boxed{\therefore A.M, \bar{x} = 41.}$$

Short cut Method :-

Here, $A = 20$

Marks (x)	Frequency (f)	$d = x - A$	fd
20	8	0	0
30	12	10	120
40	20	20	400
50	10	30	300
60	6	40	240
70	4	50	200
	N = 60		$\Sigma fd = 1260$

$$A.M.\ \bar{x} = A + \frac{\Sigma fd}{N}$$

$$= 20 + \frac{1260}{6}$$

$$= 20 + 21$$

$$\bar{x} = 41$$

$$\boxed{\therefore A.M, \bar{x} = 41.}$$

Arithmetic Mean for continuous series :-

Direct Method :-

$$\boxed{A.M.\ \bar{x} = \frac{\Sigma fm}{\Sigma f}}$$

(i) Find mid value of each class interval. i.e., 'm'.

$$\text{midvalue} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

(ii) Multiply each midvalue of the class by its frequency i.e., fm.

(iii) Sum all the multiply of each midvalue and frequency i.e., Σfm.

(iv) Sum all the frequencies of the class intervals. i.e., Σf (or) N.

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

Shortcut Method :-

(i) Find midvalue of each class interval i.e., 'm'.

$$\text{Midvalue} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

(ii) Assume one value in the midvalues sequency i.e., 'A'.

(iii) Calculate each midvalue deviation i.e., $d = m - A$

Multiply each deviation value with it's frequency. i.e, fd

~~M~~ Sum all the fd values i.e., Σfd.

Sum all the frequencies i.e, Σf

$$\boxed{A.M, \bar{x} = A + \frac{\Sigma fd}{\Sigma f}}$$ where, $d = x - A$

Σf = sum of the frequencies

(or)

$$\boxed{\bar{x} = A + \frac{\Sigma fd}{N}}$$

where, $d = x - A$

N= Sum of all the frequencies.

1) Calculate Arithmetic mean from the following data.

Marks:-	0-10	10-20	20-30	30-40	40-50	50-60
Frequency:-	5	10	25	30	20	10

Sol:- Direct Method:-

Marks (x)	Frequency (f)	Midvalues (m)	fm
0-10	5	$\frac{0+10}{2} = 5$	25
10-20	10	$\frac{10+20}{2} = 15$	150
20-30	25	$\frac{20+30}{2} = 25$	625
30-40	30	$\frac{30+40}{2} = 35$	1050
40-50	20	$\frac{40+50}{2} = 45$	900
50-60	10	$\frac{50+60}{2} = 55$	550
	$\Sigma f = 100$		$\Sigma fm = 3{,}300$

$$A.M, \bar{x} = \frac{\Sigma fm}{\Sigma f}$$

$$= \frac{3{,}300}{100}$$

$$\bar{x} = 33.$$

$$\boxed{\therefore A.M., \bar{x} = 33.}$$

Short Cut Method:-

Here

$A = 5.$

Subject : Date :

Title of the test case :

Case study No. : Page No. :

x	f	M	$d = m - A$	fd
0-10	5	5	0	0
10-20	10	15	10	100
20-30	25	25	20	500
30-40	30	35	30	900
40-50	20	45	40	800
50-60	10	55	50	500
	$\Sigma f = 100$			$\Sigma fd = 2,800.$

$$A \cdot M, \bar{x} = A + \frac{\Sigma fd}{N}$$

$$= 5 + \frac{2800}{100}$$

$$= 5 + 28$$

$$\bar{x} = 33.$$

$$\boxed{\therefore A \cdot M, \bar{x} = 33.}$$

Weighted Arithmetic Mean :-

Arithmetic mean gives equal importance of all the items. But in the case of weighted arithmetic mean the relative importance of different items is not the same for this we compute the weighted arithmetic mean.

The term weighted stands for the relative importance of different items.

$$W.A.M. \boxed{\bar{x}_w = \frac{\Sigma wx}{\Sigma w}.}$$

where wx = product of both variable and their relative weights.

i) The mean height of 25 male workers in a factory is 61 inchs and the mean height of 35 female workers in the same factory is 58 inchs. Find weighted Arithmetic mean of 60 workers in the factory?

A) Male workers in a factory, $w_1 = 25$

Male workers height, $x_1 = 61$ inchs.

Female workers in a factory, $w_2 = 35$

Female workers height, $x_2 = 58$ inchs.

$$W.A.M,\ \bar{x}_{w_1 \cdot w_2} = \frac{w_1x_1 + w_2x_2}{w_1 + w_2}$$

$$= \frac{25 \times 61 + 35 \times 58}{25 + 35}$$

$$= \frac{1525 + 2030}{60}$$

$$= \frac{3555}{60}$$

$$= 59.25$$

$$\boxed{\therefore \bar{x}_{w_1w_2} = 59.25}$$

Geometric Mean :-

Geometric mean is defined as the n^{th} root of the product of 'n' items (or) values. If there are 2 items we take the square root. there are 3 items cuberoot is taken and so.........on.

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

$$\therefore G.M = \sqrt[n]{x_1 \times x_2 \times x_3 \times \ldots\ldots \times x_n}$$

When the number of item is more than '3' the task of multiplying the numbers and exactly root ~~because~~ become more difficult to calculate. For this simple calculations logarithms are use.

In Individual Series :- $G.M = A.L\left[\frac{\Sigma \log x}{N}\right]$

In Discrete Series :- $G.M = A.L\left[\frac{\Sigma f \log x}{\Sigma f}\right]$

In Continuous Series :- $G.M = A.L\left[\frac{\Sigma f \log m}{\Sigma f}\right]$

Note :- calculating of Anti logarithms i.e., A.L = Shift + log + value in scientific calculator.

Problems :-

Individual Series :-

1) Calculate Geometric Mean from the following Data.

85, 70, 15, 75, 500, 8, 45, 250, 40, 36.

Sol :-

x	$\log x$
85	1.9294
70	1.8450
15	1.1760
75	1.8750
500	2.6989
8	0.9030
45	1.6532
250	2.3979

Here,

N = 10.

40	1.6020
36	1.5563
	$\Sigma \log x = 17.6367$

$$G.M. = A.L\left[\frac{\Sigma \log x}{N}\right]$$

$$= A.L\left[\frac{17.6367}{10}\right]$$

$$= A.L[1.76367]$$

$$\boxed{\therefore G.M. = 58.04}$$

Discrete Series:

1) Calculate Geometric mean for the following data.

x:	8	9	10	11	12	13	14
f:	11	8	6	9	7	3	1

Sol:

x	f	$\log x$	$f\log x$
8	11	0.9030	9.933
9	8	0.9542	7.6336
10	6	1	6
11	9	1.0413	9.3717
12	7	1.0791	7.5537
13	3	1.1139	3.3417
14	1	1.1461	1.1461
	$\Sigma f = 45$		$\Sigma f\log x = 44.9798$

$$G.M. = A.L\left[\frac{\Sigma f\log x}{\Sigma f}\right] \Rightarrow A.L\left[\frac{44.9798}{45}\right] \Rightarrow A.L[0.9995]$$

$$\boxed{\therefore G.M = 9.9884}$$

Continuous Series :-

1) Calculate Geometric Mean for the following distribution.

class Intervals (x) :	0-10	10-20	20-30.	30-40	40-50
Frequency (f) :	5	7	15	25	8

Sol :-

class Interval(x)	Frequency (f)	Midvalues (m)	log m	flogm
0 – 10	5	$\frac{0+10}{2} = 5$	0.6989	3.4945
10 – 20	7	15	1.1760	8.232
20 – 30	15	25	1.3979	20.9685
30 – 40	25	35	1.5440	38.6
40 – 50	8	45	1.6532	13.2256
	Σf = 60			Σflogm = 84.5206.

$$\therefore G.M = A.L\left[\frac{\Sigma f\log m}{\Sigma f}\right]$$

$$= A.L\left[\frac{84.5206}{60}\right]$$

$$= A.L\left[1.4086\right]$$

$$\boxed{\therefore G.M = 25.6212.}$$

Harmonic Mean (H.M) :-

The Harmonic mean is based on the reciprocals of numbers averaged. It is defined as the reciprocal of the arithmetic mean of the individual observations.

In Individual Series :- $H.M = \frac{N}{\Sigma(1/x)}$

In Discrete series :- $H.M = \frac{N}{\Sigma(f/x)}$

In continuous Series :- $H.M = \frac{N}{\Sigma(f/m)}$

Individual Series :-

1) Find Harmonic mean for the following distribution.

2574, 475, 75, 5, 0.8, 0.08, 0.005, 0.0009.

Sol:-

x	$1/x$
2574	0.0003
475	0.0021
75	0.0133
5	0.2
0.8	1.25
0.08	12.5
0.005	200
0.0009	1111.111
	$\Sigma(1/x) = 1325.07$

N = 8.

$$H.M. = \frac{N}{\Sigma(1/x)} = \frac{8}{1325.07}$$

$$\boxed{\therefore H.M = 0.006}$$

Subject :　　　　　　　　　　Date :

Title of the test case :

Case study No. :　　　　　　　　Page No. :

Discrete Series :-

1) Calculate Harmonic mean for the following distribution

Marks :	10	20	25	40	50
No. of students :	20	30	50	15	5

Marks (x)	No. of Students (f)	f/x.
10	20	20/10 = 2
20	30	30/20 = 1.5
25	50	50/25 = 2
40	15	15/40 = 0.375
50	5	5/50 = 0.1
	N = 120.	Σ(f/x) = 5.975

$$\therefore H.M = \frac{N}{\Sigma(f/x)} = \frac{120}{5.975} = 20.0836.$$

$$\boxed{\therefore H.M = 20.0836}$$

Continuous Series :-

1) Calculate Harmonic mean from the following distribution.

Marks (x) :	10-20	20-30	30-40	40-50	50-60
Frequency (f) :	4	6	10	7	3

Marks(x)	Midvalues (m)	frequency(f)	f/m
10-20	15	4	4/15 = 0.2666
20-30	25	6	0.24
30-40	35	10	0.2857
40-50	45	7	0.1555
50-60	55	3	0.0545
		N = 30	Σ(f/m) = 1.0023

$$\therefore H.M = \frac{N}{\Sigma(f/m)}$$

$$= \frac{30}{1.0023}$$

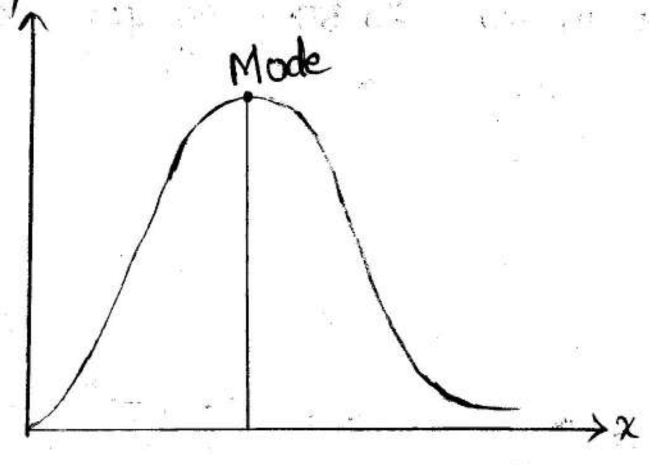
$$= 29.9311\ldots$$

$$\boxed{\therefore H.M = 29.9311}$$

MODE :-

The mode (or) model value is a value in the given series of observations which occurs the greatest frequency.

Graphically :-



Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

The value of the variable at which the curve reaches a maximum is called the "mode".

Individual Series :-

* Arrange in Ascending Order.

* Note the terms that occurring maximum number of values then the term is "mode."

Problem :-

1) Find the mode from the following data.

12, 14, 16, 18, 26, 16, 20, 16, 11, 12, 16, 15, 20, 24.

Sol:- The Ascending Order of the given data is

11, $\boxed{12, 12}$, 14, 15, $\boxed{16, 16, 16, 16}$, 18, $\boxed{20, 20}$, 24, 26.

Here, repeating terms are 12, 16, 20

12 - 2 times

16 - 4 times.

20 - 2 times.

Maximum number of repeating term is 16. i.e., 4 times.

$\boxed{\therefore \text{Mode} = 16.}$

Discrete Series :-

In discrete series, mode is known by inspection method i.e., the variable which is having highest frequency is called "Mode."

1) find mode for the following distribution.

x :	4	7	11	16	25
f :	3	9	14	21	13

Sol:

x :	4	7	11	16	25
f :	3	9	14	21	13

Highest frequency = 21

Highest frequency corresponding variable = 16.

$\therefore$ Mode = 16.

The highest frequency having the variable is 16.

$\boxed{\therefore \text{Mode} = 16.}$

In Continuous Series :-

Here we are using the formula.

$$\boxed{\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C.I.}$$

Where, L = Lower limit of the class Interval.

f_1 = frequency to the class of mode/model value.

f_0 = frequency before preceding value of the model interval.

f_2 = frequency after succeding value of the model interval.

C.I = Length of the class Interval.

Subject : Date :

Title of the test case :

Case study No. : Page No. :

1) Calculate mode from the following series.

class Interval :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency :	4	13	21	44	33	22	7

Sol:-

class Interval	Frequency
0-10	4
10-20	13
20-30	21 - f_0
L $\boxed{30}$-40	44 - f_1
40-50	33 - f_2
50-60	22
60-70	7
	N = 144

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C.I$$

$$= 30 + \frac{44-21}{2(44)-21-33} \times 10$$

$$= 30 + \frac{23}{88-54} \times 10$$

$$= 30 + \frac{23}{34} \times 10 \Rightarrow 30 + 0.6764 \times 10$$

$$\Rightarrow 30 + 6.764$$

$$\Rightarrow 36.764.$$

$$\boxed{\therefore \text{Mode} = 36.764}$$

MEDIAN :-

Median is a value that divides the series into 2 equal parts. In some cases median is the no. of terms is less than the median (or) the no. of terms is more than the median (or) equal the median. Median is standard by the following series.

Individual Series :-

* Arrange the given data in Ascending Order.

* In individual series 2 cases are existed based on the observations even (or) odd.

Case 1 :-

The number of terms in given data are odd number then we have to choose middle term is the "Median".

1) The Income of five employees are given find median for the given data.

5900, 6950, 7020, 7200, 7280.

Sol:- 5900, 6950, 7020, 7200, 7280.

Median = Middle term.

∴ Median = 7020.

Case 2 :-

The number of terms in given data are even number then we have to choose the value of median is sum of the middle two terms divided by 2.

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

1) Find Median for the following Series.

20, 68, 50, 15, 35, 98, 33, 44, 56, 64.

Sol:- Arrange in Ascending Order.

15, 20, 33, 35, 44, 50, 56, 64, 68, 96.

$$\text{Median} = \frac{N+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5$$

= 5th and 6th terms are the middle values.

$$= \frac{44+50}{2} = \frac{94}{2} = 47.$$

$$\boxed{\therefore \text{Median} = 47.}$$

Discrete Series :-

* Arrange the data in Ascending order.
* Find cummulative frequencies of the given frequencies.
* Apply the formula median = $\frac{N+1}{2}^{th}$ item.
* Now look at the cummulative frequency column and find the total which is equal to $\frac{N+1}{2}$ (or) next highest determined value of the variable to the corresponding cummulative frequency. That gives the value of median.

1) From the following data find the value of median.

Income :	4000	4500	5800	5060	6600	5380
No. of Persons :	24	26	16	20	6	30

Sol: Arrange the given data in Ascending Order.

Income (x)	No. of Persons (f)	cummulative Frequency.
4000	24	24
4500	26	50 = 24+26.
$\boxed{5060}$	20	70 = 50+20 — Median
5380	30	100 = 70+30
5800	16	116 = 100+16
6600	6	122 = 116+6
	N = 122.	

$$\text{Median} = \frac{N+1}{2}^{\text{th item}} = \frac{122+1}{2} = \frac{123}{2} = 61.5$$

$$\boxed{\therefore \text{ Median} = 5060.}$$

Continuous Series:-

* Arrange data in Ascending order.

* Calculate the cummulative frequencies.

* Apply the formula median $= L + \dfrac{\frac{N}{2} - c.f}{f} \times C.I.$

where, L = Lower limit of the median class Interval.

c.f = cummulative frequency of the preceding the value of the median class.

Subject : Date :
Title of the test case :
Case study No. : Page No. :

f = frequency of the median class.

C·I = Length of the class Interval.

1) Calculate median for the following distribution.

Marks:	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No. of Students:	7	15	24	31	42	30	26	15	10

Solution

Given,

Marks (x)	No. of Students (f)	C.f.
5-10	7	7
10-15	15	22
15-20	24	46
20-25	31	77
25-30	42	119
30-35	30	149
35-40	26	175
40-45	15	190
45-50	10	200.
	N = 200	

BALAJI INSTITUTE OF I.T. & MANAGEMENT KADAPA ENTER TO LEARN - LEAVE TO SERVE

Calculate

$\frac{N}{2} = \frac{200}{2} = 100$

$$\text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times C.I$$

$$= 25 + \frac{100-77}{42} \times 5 \Rightarrow 25 + \frac{23}{42} \times 5 \Rightarrow 25 + 2.7380$$

Prepared By:
M. Navaneeth Kumar Reddy
Asst.Prof

$\Rightarrow 27.7380 \Rightarrow 27.73$.

$\boxed{\therefore \text{Median} = 27.73.}$

Measure Of Dispersion :-

According to "A.L. Bowley", Dispersion is the measure of the variation of the items.

Properties of a good measure of Dispersion :-

* It should be simple to understand.

* It should be easy to calculate.

* It should be strongly defined.

* It should be based on each & every item of distribution.

* It should be used ~~for further~~ algebric expressions.

Methods of studying Variation :-

* Range.

* Inter quartile deviation / quartile deviation.

* Mean deviation.

* standard deviation.

* co-efficient of variation.

Range :-

Range is the simplest method of studying dispersion. It is defined as the difference between the value of the smallest item and the value of the largest item included in the distribution.

Symbolically, $\boxed{\text{Range} = L - S.}$

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

where, L = largest value
S = smallest value.

The relative measure corresponding to range called "co-efficient of range" obtained by applying the formula.

$$\boxed{\text{co-efficient of range} = \frac{L-S}{L+S}.}$$

Measure of Dispersion:-

According to A.L. Bowlany "Dispersion is the measure of the variation of the items"

According to Brooks & Dick, "Dispersion is the degree of the scatter (or) variation of the variable about a central value."

Significance of Measuring Variation:-

* To determine the repliability of an average.
* To serve as a basis for the control of the variability.
* To compare two (or) more series with regard to their variability.
* To facilitate the use of other statistical measures.

Properties of a Good Measure of Variation:-

* It should be simple to understand.
* It should be easy to compute.
* It should be strongly define.
* It should be based on each & every item of the

distribution.

- * It should be amenable to further algebraic treatment.
- * It should have sampling stability.
- * It should be unduly affected by the extreme items.

Methods of Studying Variation :-

The following are the important methods of studying variation.

- * The Range.
- * The Interquartile Range & the quartile deviation.
- * The mean deviation | Average deviation.
- * The standard deviation.
- * co-efficient of variation.

Individual Series :-

1) Find the range & co-efficient of range for the following observations.

$10, 8, 5, 10, 9, 14, 7.$

A) Largest value, $L = 14.$

Smallest value, $S = 5$

$$\text{range} = L - S$$
$$= 14 - 5$$
$$= 9.$$

$\therefore$ co-efficient of range $= \dfrac{L-S}{L+S}$

$$= \frac{14-5}{14+5}$$
$$= \frac{9}{19}$$
$$= 0.473$$

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

Discrete Series :-

1) Find the range and co-efficient of range for the following data.

x :	11	18	29	33	37	39	40	42	43
f :	100	105	91	82	61	32	70	88	67.

Sol:- Largest value, L = 43

Smallest value, S = 11

$$\text{Range} = L-S = 43-11 = 32.$$

$$\therefore \text{co-efficient of range} = \frac{L-S}{L+S} = \frac{43-11}{43+11} = \frac{32}{54} = 0.5925.$$

Continuous Series :-

1) Find range & co-efficient of range for the following data.

Marks :	10-20	20-30	30-40	40-50	50-60
No. of students :	22	28	44	43	49.

Sol:- Largest value, L = 60

Smallest value, S = 10

$$\text{Range} = L-S = 60-10 = 50.$$

$$\text{co-efficient of range} = \frac{L-S}{L+S} = \frac{60-10}{60+10} = \frac{50}{70} = 0.7142.$$

Inter quartile deviation / Quartile deviation :-

The inter quartile deviation represent the difference between the third quartile to the first quartile Symbolically, Interquartile range = $Q_3 - Q_1$. and.

quartile deviation is quartile range divided by '2.'

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2}$$

$Q_1 =$ Size of $\left(\frac{n+1}{4}\right)$ item

$Q_3 =$ Size of $3\left(\frac{n+1}{4}\right)$ item.

Individual Series :-

1) Find the quartile deviation and co-efficient of Q.D for the following data.

8, 10, 14, 22, 26, 28, 30, 36, 44, 59, 64.

Sol:- Arrange the data in Ascending order.

8, 10, 14, 22, 26, 28, 30, 36, 44, 59, 64

N = 11. (or) n = 11.

$Q_1 =$ Size of $\left(\frac{n+1}{4}\right)$

= Size of $\left(\frac{11+1}{4}\right)$

= Size of $\left(\frac{12}{4}\right)$

= Size of (3) item.

$Q_1 = 14.$

$Q_3 =$ Size of $3\left(\frac{n+1}{4}\right)$

= Size of $3\left(\frac{11+1}{4}\right)$

= Size of $3\left(\frac{12}{4}\right)$

= Size of $3(3)$

= Size of 9th item.

$Q_3 = 44.$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{44-14}{2} = \frac{30}{2} = 15.$$

$$\text{co-efficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{44-14}{44+14} = \frac{30}{58} = 0.517$$

Subject : Date :

Title of the test case :

Case study No. : Page No. :

Discrete Series :-

1) Find Q.D & co-efficient of Q.D for the following data.

x :	40	45	50	55	60	65	70	75	80
f :	20.	36	44	50	80	30	30	16	14.

Sol:-

x	f	c.f
40	20	20
45	36.	56
$\boxed{50}$	44	100
55	50	150
60	80	230
$\boxed{65}$	30	260
70	30	290
75	16	306
80	14	320
	N = 320	

$Q_1 = \text{Size of } \left(\frac{n+1}{4}\right)$

$= \left(\frac{320+1}{4}\right)$

$= \frac{321}{4}$

$Q_1 = \text{Size of } (80.25)$

$Q_1 = 50.$

$Q_3 = \text{Size of } 3\left(\frac{n+1}{4}\right)$

$= 3(80.25)$

$= 240.75$

$Q_3 = \text{Size of } (240.75)$

$Q_3 = 65$

$Q.D = \frac{Q_3 - Q_1}{2} = \frac{65-50}{2} = \frac{15}{2} = 7.5$

$\therefore$ co-efficient of $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{65-50}{65+50} = \frac{15}{115} = 0.1309$

Continuous Series :-

In continuous series, $Q_1 = L_1 + \dfrac{\frac{N}{4} - c.f_1}{f_1} \times C.I.$

$$Q_3 = L_3 + \frac{3(\frac{N}{4}) - cf_3}{f_3} \times C.I.$$

where, L_1, L_3 = lower limit of the class intervals.

N = Sum of the frequency.

c.f = cummulative frequencies preceeding value of C.I.

f_1, f_3 = frequency of C.I.

C.I = Length of class Interval.

1) Find Q.D & co-efficient of Q.D for the following data.

x:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80.
f:	5	8	7	12	28	20	10	10

Sol:-

x	f	c.f
0-10	5	5
10-20	8	13
20-30	7	20 cf_1
L_1 $\boxed{30}$- 40	12 f_1	32 cf_2
40-50	28 f_2	60 cf_3
L_3 $\boxed{50}$- 60	20 f_3	80
60-70	10	90
70-80	10	100
	N =100	

$\frac{N}{4} = \frac{100}{4} = 25.$

$\frac{N}{2} = \frac{100}{2} = 50.$

$Q_1 = L_1 + \dfrac{\frac{N}{4} - cf_1}{f_1} \times C.I.$

$= 30 + \dfrac{25-20}{12} \times 10.$

$= 30 + \dfrac{5}{12} \times 10$

$= 30 + 0.4166 \times 10$

$Q_3 = L_3 + \dfrac{3(\frac{N}{4}) - cf_3}{f_3} \times C.I.$

$= 50 + \dfrac{75-60}{20} \times 10$

$= 50 + \dfrac{15}{20} \times 10$

$= 50 + 7.5$

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

$= 30 + 4.166$

$Q_1 = 34.166.$

$Q_3 = 57.5.$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{57.5 - 34.166}{2} = \frac{23.39}{2} = 11.67$$

$$\text{co-efficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{57.5 - 34.166}{57.5 + 34.166} = \frac{23.34}{91.66}$$

$$= 0.2546$$

Mean Deviation:-

Mean deviation is also known as the average deviation. It is the difference between the items in a distribution and the median, mean (or) mode of that series

In Individual series:-

Mean deviation, $M.D = \frac{\Sigma D}{N}$ here $D = |x - \bar{x}|$

mean, $D = |x - \text{mean}|$

median, $D = |x - \text{median}|$

mode, $D = |x - \text{mode}|$

$\therefore$ co-efficient of mean deviation $= \frac{\text{Mean deviation}}{\text{Mode / Median / Mean.}}$

1) Calculate the mean deviation and co-efficient of mean deviation with the help of mean, mode, median from the following data.

4, 7, 7, 7, 9, 9, 10, 12, 15.

Sol:- **With Mean :-**

x	$D = \lvert x - \overset{\text{Mean}}{\bar{x}} \rvert$
4	$\lvert 4-9 \rvert = 5$
7	$\lvert 7-9 \rvert = 2$
7	$\lvert 7-9 \rvert = 2$
7	$\lvert 7-9 \rvert = 2$
9	$\lvert 9-9 \rvert = 0$
9	$\lvert 9-9 \rvert = 0$
10	$\lvert 10-9 \rvert = 1$
12	$\lvert 12-9 \rvert = 3$
15	$\lvert 15-9 \rvert = 6$
$\Sigma x = 80$	$\Sigma D = 21.$

Mean, $\bar{x} = \frac{\Sigma x}{N}$

$= \frac{80}{9}$

$= 8.88$

$\bar{x} \approx 9.$

$\therefore$ Mean deviation, $M.D = \frac{\Sigma D}{N}$

$= \frac{21}{9}$

$= 2.33.$

$\therefore$ co-efficient of mean deviation $= \frac{M.D}{mean} = \frac{2.33}{9} = 0.259$

With Mode :- $N = 9.$

x	$D = \lvert x - \overset{\text{Mode}}{\bar{x}} \rvert$
4	$\lvert 4-7 \rvert = 3$
7	$\lvert 7-7 \rvert = 0$
7	$\lvert 7-7 \rvert = 0.$
7	$\lvert 7-7 \rvert = 0$
9	$\lvert 9-7 \rvert = 2$
9	$\lvert 9-7 \rvert = 2$
10	$\lvert 10-7 \rvert = 3$
12	$\lvert 12-7 \rvert = 5$
15	$\lvert 15-7 \rvert = 8$
	$\Sigma D = 23.$

$\therefore$ Mode, $\bar{x} = 7$. (highest number of repeating term.)

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

$\therefore$ Mean deviation, $M.D = \frac{\Sigma D}{N}$

$= \frac{23}{9}$

$= 2.55.$

$\therefore$ co-efficient of mean deviation $= \frac{\text{Mean deviation}}{\text{Mode}}$

$= \frac{2.55}{7}$

$= 0.364.$

With Median :-

Median

x	$D=\lvert x-\bar{x}\rvert$
4	$\lvert 4-9\rvert = 5$
7	$\lvert 7-9\rvert = 2$
7	$\lvert 7-9\rvert = 2$
7	$\lvert 7-9\rvert = 2$
9	$\lvert 9-9\rvert = 0$
9	$\lvert 9-9\rvert = 0$
10	$\lvert 10-9\rvert = 1$
12	$\lvert 12-9\rvert = 3$
15	$\lvert 15-9\rvert = 6$
	$\Sigma D = 21.$

$\therefore$ Median = 9, (here no. of terms is 9. So, odd terms existed. So, middle term is called "Median".)

$\therefore$ Mean deviation, $M.D = \frac{\Sigma D}{N} = \frac{21}{9} = 2.33.$

$\therefore$ co-efficient of mean deviation $= \frac{M.D}{\text{Median}}$

$= \frac{2.33}{9}$

$= 0.259$

Discrete Series:-

Mean deviation $= \frac{\Sigma FD}{N}$

where, f= frequency

D = $|x-\bar{x}|$ ↳ mean, median, mode

N = Sum of the frequency.

∴ co-efficient of mean deviation = $\frac{\text{Mean deviation}}{\text{Mean} | \text{Mode} | \text{Median}}$.

1) Calculate Mean deviation and co-efficient of mean deviation for the following data with the help of mean, mode, median?

x :	10	11	12	13	14
f :	3	12	18	12	3.

Sol:- **With Mean:-**

x	f	fx	$D = \lvert x-\bar{x}\rvert$	fD
10	3	30	$\lvert 10-12\rvert = 2$	$3\times2 = 6$
11	12	132	$\lvert 11-12\rvert = 1$	$12\times1 = 12$
12	18	216	$\lvert 12-12\rvert = 0$	$18\times0 = 0$
13	12	156	$\lvert 13-12\rvert = 1$	$12\times1 = 12$
14	3	42	$\lvert 14-12\rvert = 2$	$14\times2 = 28$
	N=48	$\Sigma fx = 576$		$\Sigma fD = 56$

Mean, $\bar{x} = \frac{\Sigma fx}{N}$

$= \frac{576}{48}$

$\bar{x} = 12$

∴ Mean deviation, M.D $= \frac{\Sigma fd}{N} = \frac{56}{48} = 1.1666.$

∴ co-efficient of mean deviation $= \frac{M.D}{\text{mean}}$

$= \frac{1.1666}{12}$

$= 0.0972$.

Subject : Date :

Title of the test case :

Case study No. : Page No. :

With Mode :-

x	f	$D=\lvert x-\bar{x}\rvert$	fD
10	3	$\lvert 10-12\rvert=2$	6
11	12	$\lvert 11-12\rvert=1$	12
12-mode	18	$\lvert 12-12\rvert=0$	0
13	12	$\lvert 13-12\rvert=1$	12
14	3	$\lvert 14-12\rvert=2$	6
	$N=48$		$\Sigma fD=36$

∴ Mode = highest frequency variable $=12$.

∴ Mean deviation, $M{\cdot}D = \frac{\Sigma fd}{N} = \frac{36}{48} = 0{\cdot}75$.

∴ co-efficient of $M{\cdot}D = \frac{\text{Mean deviation}}{\text{Mode}} = \frac{0{\cdot}75}{12} = 0{\cdot}06$

With Median :-

x	f	cf	$d=\lvert x-\bar{x}\rvert$	fd
10	3	3	$\lvert 10-12\rvert=2$	6
11	12	15	$\lvert 11-12\rvert=1$	12
12	18	33	$\lvert 12-12\rvert=0$	0
13	12	45	$\lvert 13-12\rvert=1$	12
14	3	48	$\lvert 14-12\rvert=2$	6
				$\Sigma fd=36$

calculate $\frac{N+1}{2}$

$= \frac{48+1}{2}$

$= \frac{49}{2}$

$= 24{\cdot}5$.

∴ Median $\bar{x} = 12$.

Mean deviation, $M{\cdot}D = \frac{\Sigma fD}{N} = \frac{36}{48} = 0{\cdot}75$.

∴ co-efficient of $M{\cdot}D = \frac{\text{Mean deviation}}{\text{Median}} = \frac{0{\cdot}75}{12} = 0{\cdot}06$

Continuous Series:-

Mean deviation = $\frac{\Sigma fd}{N}$

where, d= deviation $|x-\bar{x}|$

$\bar{x}$ = mean/mode/median.

N= Sum of the frequency.

co-efficient of M·D = $\frac{M \cdot D}{Mean/Mode/Median}$.

1) calculate mean deviation and co-efficient of mean deviation from the following data with the help of mean?

class (x):	0-10	10-20	20-30	30-40	40-50	50-60.
Frequency(f):	5	8	7	12	28	20

Sol:-

x	f	m	fm	$d=\lvert x-\bar{x}\rvert$	fd
0-10	5	5	25	34	170
10-20	8	15	120	24	192
20-30	7	25	175	14	98
30-40	12	35	420	4	48
40-50	28	45	1260	6	168
50-60	20	55	1100	16	320
	N=80		$\Sigma fm=3100$		$\Sigma fd=996$.

Mean, $\bar{x} = \frac{\Sigma fm}{N} = \frac{3100}{80} = 38.75$

$\boxed{\bar{x} \simeq 39.}$

Mean deviation, $M \cdot D = \frac{\Sigma fd}{N} = \frac{996}{80} = 12.45$.

∴ co-efficient of Mean deviation = $\frac{\text{Mean deviation}}{\text{Mean}}$

$= \frac{12.45}{39} = 0.3192$

Subject :	Date :
Title of the test case :
Case study No. :	Page No. :

Standard Deviation :-

The standard deviation concept was introduced by "Karl Pearson" in the year 1823. It is mostly used to measure the studying of dispersion. standard deviation is also known as "Root mean square deviation". For this reason standard deviation is square root of the mean square deviation from the arithmetic mean. Symbolically, it is denoted by "σ".

$$S.D, \sigma = \sqrt{\frac{\Sigma x^2}{N}}$$

Note :- The relative measure of standard deviation is called "Variance". Symbolically it is denoted by "σ^2".

In Individual Series :-

$$S.D, \sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

where, d = deviation calculated from mean $\bar{x}$

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

N = Total no. of variables.

Variance $= (\sigma)^2$.

1) Calculate standard deviation and variance from the following data.

120, 100, 160, 100, 220, 130, 150, 170, 150, 200.

Sol:- N = 10.

x	$d = (x - \bar{x})$	d^2
120	120-150 = -30	$(-30)^2 = 900$
100	100-150 = -50	$(-50)^2 = 2500$
160	160-150 = 10	$(10)^2 = 100$
100	100-150 = -50	$(-50)^2 = 2500$.
220	220-150 = 70	$(70)^2 = 4900$
130	130-150 = -20	$(-20)^2 = 400$
150	150-150 = 0	$(0)^2 = 0$
170	170-150 = 20	$(20)^2 = 400$
150	150-150 = 0	$(0)^2 = 0$
200	200-150 = 50	$(50)^2 = 2500$
$\Sigma x = 1500$	$\Sigma d = 0$.	$\Sigma d^2 = 14,200$

Mean $\bar{x} = \frac{\Sigma x}{N}$

$= \frac{1500}{10}$

$\bar{x} = 150$.

standard deviation, $\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} = \sqrt{\frac{14200}{10} - \left(\frac{0}{10}\right)^2}$

$= \sqrt{1420 - 0} = \sqrt{1420} = 37.68$

$\sigma = 37.68$.

Variance $= (\sigma)^2 = (37.68)^2 = 1420$.

$\sigma^2 = 1420$

Discrete Series :-

standard deviation, $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

fd = frequency with deviation.

N = sum of the frequencies.

1) Calculate S.D and variance from the following Data.

x:	10	20	30	40	50	60
f:	8	12	20	10	7	3

Subject : Date :

Title of the test case :

Case study No. : Page No. :

Sol:-

x	f	fx	$d\ (x-\bar{x})$ 31	d^2	fd	fd^2
10	8	80	-21	441	-168	3528
20	12	240	-11	121	-132	1452
30	20	600	-1	1	-20	20
40	10	400	9	81	90	810
50	7	350	19	361	133	2527
60	3	180	29	841	87	2523
	$N=60$	$\Sigma fx=1850$			$\Sigma fd=-10.$	$\Sigma fd^2=10{,}860.$

Mean, $\bar{x} = \frac{\Sigma fx}{N} = \frac{1850}{60} = 30.8$

$$\boxed{\bar{x} \approx 31.}$$

standard deviation, S.D, $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

$$= \sqrt{\frac{10860}{60} - \left(\frac{-10}{60}\right)^2}$$

$$= \sqrt{181 - (0.166)^2}$$

$$= \sqrt{181 - 0.027}$$

$$= \sqrt{180.973}$$

$$\sigma = 13.45.$$

· Variance, $\sigma^2 = (13.45)^2 = 180.973.$

II) Continuous Series :-

standard deviation, S.D, $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C.$

where, fd = frequency with mid values deviation.

c = length of the class Interval.

1) Calculate standard deviation and variance from the following data.

class (x):	0-10	10-20	20-30	30-40	40-50	50-60
Frequency(f):	8	12	20	10	7	3

Sol:-

x	f	m	fm	$d(x-\bar{x})$ $\bar{x}=26$	d^2	fd	fd^2
0-10	8	5	40	−21	441	−168	3528
10-20	12	15	180	−11	121	−132	1452
20-30	20	25	500	−1	1	−20	20
30-40	10	35	350	9	81	90	$8\overset{1}{\cancel{4}}0$
40-50	7	45	315	19	361	133	2527
50-60	3	55	165	29	841	87	2523
	N=60		$\Sigma fm=1550$			$\Sigma fd=-10$	$\Sigma fd^2=10860.$

$$\text{Mean}, \bar{x} = \frac{\Sigma fm}{N} = \frac{1550}{60} = 25.83$$

$$\boxed{\therefore \bar{x} \approx 26.}$$

- Standard deviation, S.D, $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$

$$= \sqrt{\frac{10860}{60} - \left(\frac{-10}{60}\right)^2} \times 10$$

$$= \sqrt{181 - (0.166)^2} \times 10 = \sqrt{181 - 0.027} \times 10$$

$$= \sqrt{180.973} \times 10$$

$$= 13.45 \times 10$$

$$\sigma = 134.5.$$

$$\text{Variance} = \sigma^2 = (134.5)^2$$

$$\sigma^2 = 18090.25.$$

Co-efficient of Variation :-

Standard deviation is discussed absolute measure of dispersion the corresponding relative measure is called "co-efficient of Variation". This method is developed by "Karl Pearson." This is mostly used to calculate the relative measure is called "Variation". It is used in such problems where we want to compare the variability of two (or) more series.

$$\therefore \text{co-efficient of variation} = \frac{\sigma}{\bar{x}} \times 100.$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$$

where, σ = standard deviation.
$\bar{x}$ = mean.

1) From the prices of shares of 'x' and 'y' given below. Find out which is more stable in value.

x : 35 54 52 53 56 58 52 50 51 49
y : 108 107 105 105 106 107 104 103 104 101.

Sol:-

x	$X(x-\bar{x})$	X^2	y	$Y(y-\bar{y})$	Y^2
35	-16	256	108	3	9
54	3	9	107	2	4
52	1	1	105	0	0
53	2	4	105	0	0
56	5	25	106	1	1
58	7	49	107	2	4
52	1	1	104	-1	1
50	-1	1	103	-2	4
51	0	0	104	1	1
49	-2	4	101	-4	16
$\Sigma x = 510$		$\Sigma X^2 = 350$	$\Sigma y = 1050$		$\Sigma Y^2 = 40$.

Mean, $\bar{x} = \frac{\Sigma x}{N} = \frac{510}{10} = 51.$

$\boxed{\bar{x} = 51.}$

Mean, $\bar{y} = \frac{\Sigma y}{N} = \frac{1050}{10} = 105$

$\boxed{\bar{y} = 105.}$

co-efficient of variation in $x = \frac{\sigma}{\bar{x}} \times 100$

$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{350}{10}} = \sqrt{35} = 5.91.$

co-efficient of variation, $x = \frac{\sigma}{\bar{x}} \times 100$

$= \frac{5.91}{51} \times 100 = 0.1158 \times 100$

$= 11.58.$

co-efficient of variation in $y = \frac{\sigma}{\bar{y}} \times 100$

$\sigma = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{40}{10}} = \sqrt{4} = 2.$

co-efficient of variation, $y = \frac{\sigma}{\bar{y}} \times 100 = \frac{2}{105} \times 100 = 0.0190 \times 100$

$= 1.90.$

∴ Here, co-efficient of variation in 'x' is more when compare to co-efficient of variation in 'y'. So, shares 'y' is more stable to shares 'x'.

Applications of Measure of Central Tendency & Dispersion:- Central Tendency and dispersion can be used for

* In finance measures of central tendency and dispersion is used as an indicator of the risk involved in an investment. Since, it measures the variability of returns around the expected return from an investment.

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

* Financial managers can also use expected value and dispersion action to make important inferences from the past data.

* Measures of central Tendency & dispersion is used in determining the variability in sales & earnings.

* The expected returns & its associated standard deviations are used as measure of risk on investment analysis. These measures used in comparing investments such as stocks, bonds & even mutual funds.

* The measures of central tendency and dispersion can be used to analyse sample market survey data, rates of return on a stock and economic data.

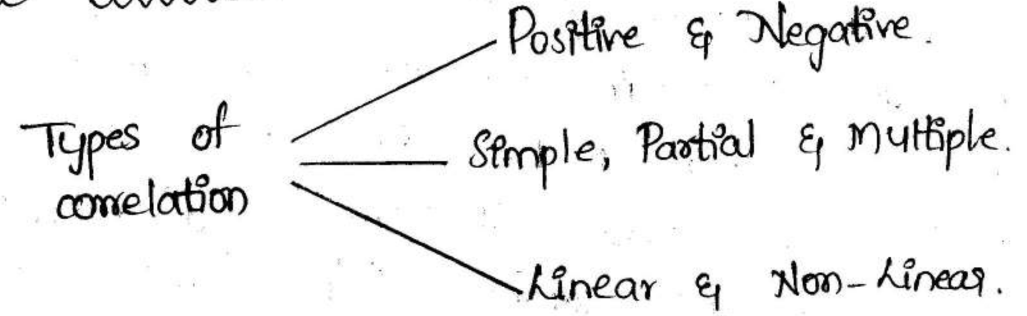
UNIT-2

CORRELATION.

Types of Correlation :-

Definition :- A statistical tool used to measure the relationship between two (or) more variables such that the movement in one variable is accompained by the movement of another is called as "Correlation."

Types of Correlation :-


Types of correlation
Positive & Negative.
Simple, Partial & Multiple.
Linear & Non-Linear.

Positive & Negative Correlation :-

Whether the correlation between the variables is positive (or) negative depends on it's direction of change. The correlation is positive when both the variables move in the same direction, i.e., when one variable increases the other on an average also increases and if one variable, decreases the other also decreases. the correlation is said to be negative when both the variables move in the opposite direction, i.e., when one variable increases the other decreases & vice versa.

Simple, Partial And Multiple Correlations :-

Whether the correlation is Simple, partial (or) multiple depends on the number of variables studied.

The correlation is said to be simple when only two variables are studied. The correlation is either multiple or partial when three (or) more variables are studied. the correlation is said to be Multiple when three variables are studied simultaneously. such as, if we want to study the relationship between the yield of wheat per acre and the amount of fertilizers and rainfall used, then it is a problem of multiple correlations.

Whereas, in the case of a partial correlation we study more than two variables, but consider only two among them that would be influencing each other such that the effect of the other influencing variable is kept constant. Such as, in the above example, if we study the relationship between the yield and fertilizers used during the periods when certain average temperature existed, then it is a problem of partial correlation.

Linear & Non-Linear (curvilinear) Correlation:-

Whether the correlation between the variables is linear (or) non-linear depends on the constancy of ratio of change between the variables. The correlation is said to be linear when the amount of change in one variable to the amount of change in another variable tends to bear a constant ratio. For example, from the values of two variables given below, it is clear that the ratio of change between the variables is the Same :

X : 10 20 30 40 50.

Y : 20 40 60 80 100.

The correlation is called as "Non-Linear (or) curvilinear" when the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable. For example, if the amount of fertilizers is doubled the yield of wheat would not be necessarily be doubled.

Thus, there are three most important types of correlation classified on the basis of movement, number and the ratio of change between the variables. The researcher must study these carefully to determine the correlation methods to be used to identify the extent to which the variables are correlated.

Methods of Determining Correlation :-

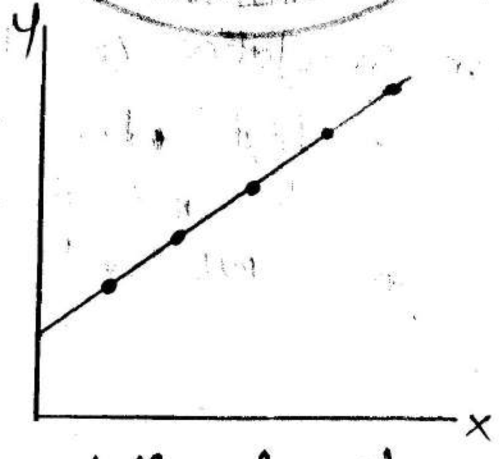
Definition :- The scatter diagram method is the simplest method to study the correlation between two variables wherein the values for each pair of a variable is plotted on a graph in the form of dots thereby obtaining as many points as the number of observations. then by looking at the scatter of several points, the degree of correlation is ascertained.

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

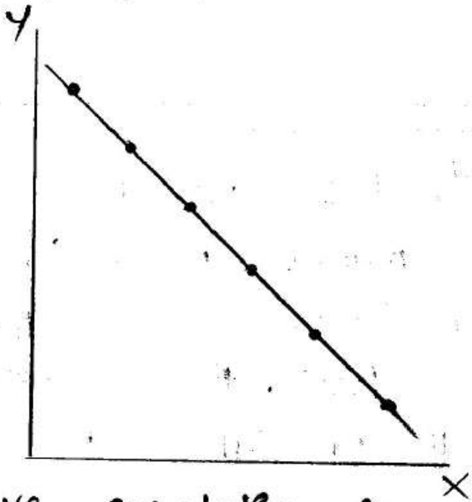
The degree to which the variables are related to each other depends on the manner in which the points are scattered over the chart. The more the points plotted are scattered over the chart, the lesser is the degree of correlation between the variables. The more the points plotted are closer to the line, the higher is the degree of correlation. The degree of correlation is denoted by "r".

The following types of scatter diagrams tell about the degree of correlation between variable x and variable y.

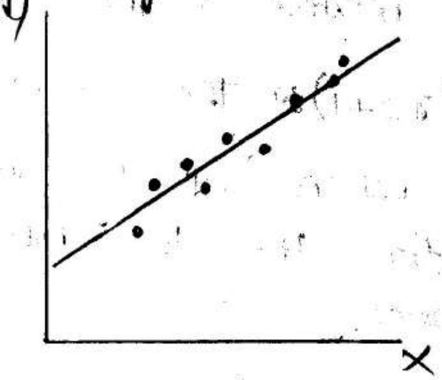
Positive Correlation ($r=+1$):- The correlation is said to be perfectly positive when all the points lie on the straight line rising from the lower left-hand corner to the upper right-hand corner.



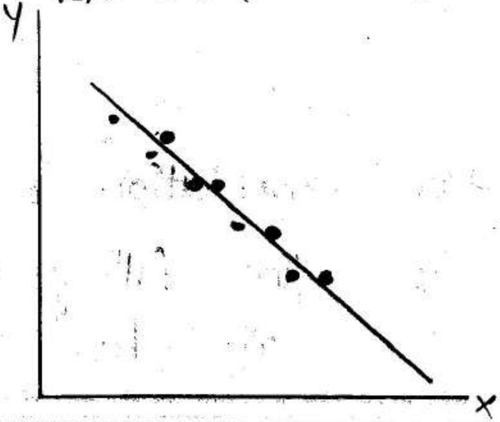
Perfect Negative Correlation ($r=-1$):- When all the points lie on a straight line falling from the upper left-hand corner to the lower right-hand corner, the variables are said to be negatively correlated.



High Degree of +ve correlation (r = +high) :- The degree of correlation is high when the points plotted fall under the narrow band and is said to be positive when these show the rising tendency from the lower left-hand corner to the upper right-hand corner.



High degree of -ve correlation (r = -high) :- The degree of negative correlation is high when the point plotted fall in the narrow band and show the declining tendency from the upper left-hand corner to the lower right-hand corner.



Subject :

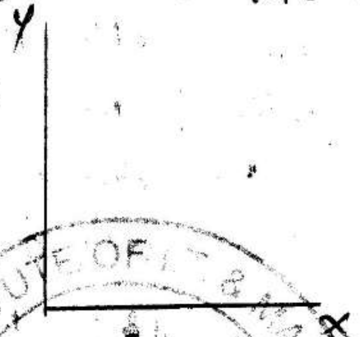
Date :

Title of the test case :

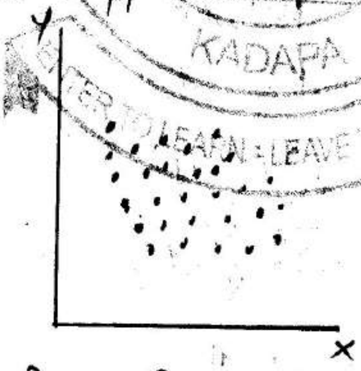
Case study No. :

Page No. :

Low Degree of +ve Corelation (r = +Low) :- The correlation between the variables is said to be low but positive when the points are highly scattered over the graph & show a rising tendency from the lower left-hand corner to the upper right-hand corner.



Low Degree of -ve Corelation (r = +Low) :- The degree of correlation is low and negative when the points are scattered over the graph and the show the falling tendency from the upper left-hand corner to the lower right-hand corner.



No Corelation (r=0) :- The variable is said to be unrelated when the points are haphazardly scattered over the graph and do not show any specific pattern. Here the correlation is absent and hence r=0.

Thus, the scatter diagram method is the simplest device to study the degree of relationship between the variables by plotting the dots of each pair of variable values given. The chart on which the dots are plotted is also called as a "Dotogram."

Karl Pearson's Co-efficient of Correlation :-

Definition :- Karl pearson's co-efficient of correlation is widely used mathematical method where in the numerical expression is used to calculate the degree and direction of the relationship between linear related variables.

Pearson's method, popularly known as a "Pearson Co-efficient of Corelation," is the most extensively used quantitative methods in practice. The co-efficient of correlation is denoted by "r".

If the relationship between two variables x and y is to be ascertained, then the following formula is used:

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

where, $\bar{x}$ = mean of x variable

$\bar{y}$ = mean of y variable.

Properties of Co-efficient of Correlation :-

* The value of the co-efficient of correlation (r) always lies between ±1. such as:

$r=+1$, perfect positive correlation.

$r=-1$, perfect negative correlation.

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

$r=0$, no correlation.

* The co-efficient of correlation is independent of the origin & scale. By origin, it means subtracting any non-zero constant from the given value of x and y the value of "r" remains unchanged. By scale it means, there is no effect on the value of "r" if the value of x and y is divided (or) multiplied by any constant.

* The co-efficient of correlation is a geometric mean of two regression co-efficient. Symbolically, it is represented as

$$r = \sqrt{b_{xy} + b_{yx}}$$

* co-efficient of correlation is "zero" when the variables x and y are independent. But, however, the converse is not true.

Assumptions of Karl Pearson's Co-efficient of Correlation:-

* The relationship between the variables is "linear", which means when the two variables are plotted, a straight line is formed by the points plotted.

* There are a large no. of independent causes that affect the variables under study so as to form a Normal Distribution. such as, variables like price, demand, supply, etc. are affected by such factors that the normal distribution is formed.

* The variables are independent of each other.

Note:- The co-efficient of correlation measures not only the magnitude of correlation but also tells the direction, such as, $r = -0.67$, which shows correlation is negative because the sign is "–" and the magnitude is 0.67.

Spearman's Rank Correlation Co-efficient:-

Definition:- The Spearman's Rank Correlation Co-efficient is the non-parametric statistical measure used to study the strength of association between the two ranked variables. This method is applied to the ordinal set of numbers, which can be arranged in order, i.e., one after the other so that ranks can be given to each.

In the rank correlation co-efficient method, the ranks are given to each individual on the basis of it's quality (or) quantity, such as ranking starts from position 1st and goes till Nth position for the one ranked last in the group.

The formula to calculate the rank correlation co-efficient is:

$$R = \frac{(1-6\Sigma D^2)}{N(N^2-1)} = \frac{(1-6\Sigma D^2)}{N^3-N}$$

where, R = Rank co-efficient of correlation.
D = Difference of ranks.
N = Number of observations.

The value of R lies between ±1 such as:

R = +1, there is a complete agreement in the order of ranks and move in the same direction.

R = −1, there is a complete agreement in the order of ranks, but are in opposite directions.

R=0, there is no association in the ranks.

While solving for the rank correlation co-efficient one may come across the following problems:

→ Where actual Ranks are given.

→ where ranks are not given.

→ Equal ranks (or) Tie in Ranks.

Where actual ranks are given:- An individual must follow the following steps to calculate the correlation coefficient:

* First, the difference between the ranks (R1-R2) must be calculated, denoted by D.
* Then, square these differences to remove the negative sign & obtain its sum ΣD^2.
* Apply the formula as shown above.

Where ranks are not given:- In case the ranks are not given, then the individual may assign the rank by taking either the highest value (or) the lowest value as 1. Whatever criteria is being decided the same method should be applied to all the variables.

Equal Ranks (or) Tie in Ranks:- In case the same ranks are assigned to two (or) more entities, then the ranks are assigned on an average basis. such as if two Individuals are ranked equal at third position, then the ranks shall be calculated as :

$(3+4)/2 = 3.5$

The formula to calculate the rank correlation co-efficient when there is a tie in the ranks is :-

$$R = 1 - \frac{6[6\Sigma d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + ----)}{N^3 - N}.$$

where m = number of items whose ranks are common.

Note:- The Spearman's rank correlation co-efficient method is applied only when the initial data are in the form of ranks, and N (Number of observations) is fairly small, i.e., not greater than 25 (or) 30.

Key differences between Correlation and Regression:-

The points given below, explains the difference between correlation & regression in detail:

* A statistical measure which determines the co-relationship (or) association of two quantities is known as "Correlation". Regression describes how an independent variable is numerically related to the dependent variable.
* Correlation is used to represent the linear relationship between two variables. On the contrary, regression is used to fit the best line and estimate one variable on the basis of another variable.

Subject :

Date :

Title of the test case :

Case study No. :

Page No. :

* In correlation, there is no difference between dependent & independent variables i.e., correlation between x and y is similar to y and x. conversely, the regression of y on x is different from x on y.

* Correlation indicates the strength of association between variables. As opposed to, regression reflects the impact of the unit change in the independent variable on the dependent variable.

* Correlation aims at finding a numerical value that expresses the relationship between variables. Unlike regression whose goal is to predict values of the random variable on the basis of the values of fixed variable.

Meaning of Regression coefficient :-

Regression co-efficient is a statistical measure of the average functional relationship between two (or) more variables. In regression analysis, one variable is considered as dependent and others as independent. Thus, it measures the degree of dependence of one variable on the others. Regression co-efficient was first used for estimating the relationship between the heights of fathers and their sons.

Properties of Regression Co-efficient :-

The important properties of regression co-efficient are given below :

* It is denoted by b.

* It is expressed in terms of original unit of data.

* Between two variables (say x and y), two values of regression co-efficient can be obtained. One will be obtained when we consider x as independent and y as dependent and the other when we consider y as independent and x as dependent. The regression co-efficient of y on x is represented as byx and that of x on y as bxy.

* Both regression co-efficients must have the same sign. If byx is positive, bxy will also be positive & viceversa.

* If one regression co-efficient is greater than unity, then the other regression co-efficient must be lesser than unity.

* The geometric mean between two regression co-efficients is equal to the co-efficient of correlation, r = doubt

* Arithmetic mean of both regression co-efficients is equal to (or) greater than co-efficient of correlation.

$(byx + bxy)/2$ = equal (or) greater than r.

Regression co-efficients are classified as :

1) Simple, partial and multiple.

2) Positive and negative and

3) Linear and non-linear.

Computation of Regression Co-efficient :-

Regression co-efficient can be worked out from both un-replicated and replicated data. For calculation of regression co-efficient from un-replicated data three estimates, viz., (1) sum of all obeervations on x and y ($\Sigma x, \Sigma y$) variables, (2) their sum of squares [Σx^2 and Σy^2] and (3) sum of products of all observations on x and y variables (Σxy).

Then regression co-efficient can be worked out as follows:

$byx = \Sigma xy - (\Sigma x . \Sigma y) / \Sigma y^2 - (\Sigma y)^2$

$bxy = \Sigma xy - (\Sigma x . \Sigma y) / \Sigma x^2 - (\Sigma x)^2$

In case of replicated data, first analysis of variances and co-variances is performed and then regression co-efficient is worked out as given below:

$byx = cov. (xy) / vx$, and $bxy = cov. (xy) / vy$.

where, cov = co-variance between x and y

vx = variance of x.

vy = variance of y.

The significance of regression co-efficient is generally tested with the help of t-test.

First t is worked out as given below:

$t = byx / SE(b)$.

The calculated value of t is compared with the table value of t at desired level of significance and appropriate degrees of freedom. If the calculated value of t is greater than table value, it is considered significant and vice versa.

The value of dependent variable can be predicated with the value of independent variable. By substitution the value of dependent variable we can get value of independent variable.

Applications of Regression Co-efficient in Genetics:-

Regression analysis has wide applications in the field of genetics and breeding as given below:

* It helps in finding out a cause and effect relationship. between two or more plant characters.
* It is useful in determining the important yield contributing characters.
* It helps in the selection of elite genotypes by indirect selection for yield through independent characters.
* It also helps in predicting the performance of selected plants in the next generation.

Properties of Regression co-efficient and regression lines:-

i) The regression co-efficients remain unchanged due to a shift of origin but change due to a shift of scale.

This property states that if the original pair of variables is (x,y) and if they are changed to the pair (u,v) where

$$u=\frac{xa}{p} \text{ and } v=\frac{yc}{q}$$

$$b_{yx}=\frac{q}{p}\times b_{vu}$$

$$\text{and } b_{xy}=\frac{p}{q}\times b_{uv}.$$

Subject	:	Date	:
Title of the test case	:		
Case study No.	:	Page No.	:

(ii) The two lines of regression intersect at the point

(Mean of "x", mean of "y"),

where x and y are the variables under consideration.

(iii) The co-efficient of correlation between two variables x & y in the simple geometric mean of the two regression co-efficients. The sign of the correlation co-efficient would be the common sign of the two regression co-efficients.

This property says that if the two regression coefficients are denoted by byx and bxy then the co-efficient of correlation is given by

$$r = \pm\sqrt{byx \times bxy}$$

If both the regression co-efficients are negative, r would be negative and if both are positive, r would assume a positive value.

(iv) The two lines of regression coincide i.e., become identical when r = -1 (or) 1 (or) in other words, there is a perfect negative (or) positive correlation between the two variables under discussion.

(v) The two lines of regression are perpendicular to each other when r=0.

Co-efficient of Corelation :-

With mean :-

* Find the mean value for the given variables x and y i.e., $\bar{x}$ & $\bar{y}$
* calculate the deviations of $X (x-\bar{x})$ and $Y (y-\bar{y})$
* Square the deviations of 'x' and 'y' series.
* Multiply the single deviation in 'x' series with single deviation in 'y' series i.e., XY.
* Apply the formula $\gamma = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 \cdot \Sigma Y^2}}$.

1) Find the co-efficient of correlation from the data marks in accounting & marks in statistics.

Marks in Accounting :- 48 35 17 23 47

Marks in statistics :- 45 20 40 35 45

Sol:- Let us consider, Marks in accounting as 'x'.
Marks in statistics as 'y'.

x	y	X	Y	X^2	Y^2	XY
48	45	48-34 14	45-37 8	196	64	112
35	20	1	-17	1	289	-17
17	40	-17	3	289	9	-51
23	35	-11	-2	-121	4	22
47	45	13	8	169	64	104
$\Sigma x = 170$	$\Sigma y = 185$			$\Sigma X^2 = 776$	$\Sigma Y^2 = 430$	$\Sigma XY = 170$

$\therefore \bar{x} = \frac{\Sigma x}{N} = \frac{170}{5} = 34.$

$\bar{y} = \frac{\Sigma y}{N} = \frac{185}{5} = 37.$

∴ co-efficient of correlation $(r) = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$

$$r = \frac{170}{\sqrt{776 \times 430}}$$

$$r = \frac{170}{\sqrt{333680}} = \frac{170}{577.6} = 0.294$$

$$\boxed{\therefore r = 0.294.}$$

Conclusion:- Here, $r>0$, then the correlation is said to be positive correlation and the variables are positively correlated.

∴ The range of the correlation is $1 > r > 0$.

Without Mean calculate the co-efficient of correlation:-

$$r = \frac{N\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

where N = Total no. of observations

xy = Product of series 'x' and series 'y'

x^2 = Square the variables in series 'x'.

y^2 = Square the variables in series 'y'.

1) calculate the coefficient of correlation from the following data.

x : 2 3 4 5 6

y : 7 9 10 14 15

Sol^n: Here, N = 5.

x	y	x^2	y^2	xy
2	7	4	49	14
3	9	9	81	27
4	10	16	100	40
5	14	25	196	70
6	15	36	225	90
$\Sigma x = 20$	$\Sigma y = 55$	$\Sigma x^2 = 90$	$\Sigma y^2 = 651$	$\Sigma xy = 241$

$$r = \frac{N\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{5 \times 241 - 20 \times 55}{\sqrt{5 \times 90 - (20)^2} \cdot \sqrt{5 \times 651 - (55)^2}}$$

$$= \frac{1205 - 1100}{\sqrt{450 - 400} \cdot \sqrt{3255 - 3025}}$$

$$= \frac{105}{\sqrt{50} \cdot \sqrt{230}} = \frac{105}{7.07 \times 15.16} = \frac{105}{107.18}$$

$$\boxed{r = 0.97.}$$

Conclusion: Here $r > 0$, the co-efficient of correlation is said to be positive & valiables are positively correlated.

Subject : Date :

Title of the test case :

Case study No. : Page No. :

Co-efficient of correlation with the help of Assumed Mean:-

co-efficient of correlation, $r = \dfrac{N\Sigma dxdy - \Sigma dx \cdot \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \cdot \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$

where, N = Total no. of observations.

dx, dy = deviations of variables in series 'x' & 'y', i.e.,

$$dx = x - A$$
$$dy = y - A$$

where, A = Assumed mean in the given series.

dxdy = product of deviation variables in series 'x' & 'y'.

dx^2, dy^2 = Squaring the deviations of series 'x' & 'y'.

1) calculate the co-efficient of correlation from the following data.

x : 2 3 4 5 6

y : 7 9 10 14 15

Sol:-

A = 2 in series x

A = 7 in series y.

x	y	dx (x-A) x-2	dy(y-A) y-7	dx^2	dy^2	dx·dy
2	7	0	0	0	0	0
3	9	1	2	1	4	2
4	10	2	3	4	9	6
5	14	3	7	9	49	21
6	15	4	8	16	64	32
		$\Sigma dx = 10$	$\Sigma dy = 20$	$\Sigma dx^2 = 30$	$\Sigma dy^2 = 126$	$\Sigma dx \cdot dy = 61$.

$$r = \frac{N\Sigma dxdy - \Sigma dx \cdot \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \cdot \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$\therefore \Sigma dx = 10, \Sigma dy = 20, \Sigma dx^2 = 30, \Sigma dy^2 = 126, \Sigma dxdy = 61.$

$$= \frac{5 \times 61 - 10 \times 20}{\sqrt{5 \times 30 - (10)^2} \cdot \sqrt{5 \times 126 - (20)^2}}$$

$$= \frac{305 - 200}{\sqrt{150 - 100} \cdot \sqrt{630 - 400}}$$

$$= \frac{105}{\sqrt{50} \cdot \sqrt{230}} = \frac{105}{7.07 \times 15.16} = \frac{105}{107.18}$$

$$\boxed{\therefore r = 0.97}$$

$r > 0$.

Conclusion:- Here. $r > 0$, the co-efficient of correlation is said to be positive & the variables are positively correlated.

Rank correlation/ Spearsman Correlation:-

In 1940's charles edwards & Pearson proposed a method for the purpose of calculated the rank correlation. This is the simplest method.

$$r_k = 1 - \frac{6\Sigma d^2}{N^3 - N}$$

where N = No. of items

d = difference between ranks.

In rank correlation 3 situations are involved:

- * when the ranks are given.
- * when the ranks are not given.
- * when the ranks are equal.

When the ranks are given:-

$$\text{formula, } r_k = 1 - \frac{6\Sigma d^2}{N^3 - N}$$

where, d= deviation between the ranks i.e., $R_x - R_y$.

N = Total no, of observations.

This is the situation when the ranks are given by the examiner.

1) Two Ladies were asked the rank and different types of lipsticks the rank given by them are as follows:

Lipsticks:	A	B	C	D	E	F	G
Neelu (R_1):	2	1	4	3	5	7	6
Neena(R_2):	1	3	2	4	5	6	7

Calculate spearsman rank correlation co-efficient?

Sol:- N=7.

Lipsticks	Neelu (R_1)	Neena (R_2)	$d(R_1 - R_2)$	d^2
A	2	1	1	1
B	1	3	-2	4
C	4	2	2	4
D	3	4	-1	1
E	5	5	0	0
F	7	6	1	1
G	6	7	-1	1
				$\Sigma d^2 = 12.$

∴ Rank correlation, $r_K = 1 - \frac{6\Sigma d^2}{N^3-N}$

$$= 1 - \frac{6(12)}{7^3-7} = 1 - \frac{72}{343-7} = 1 - \frac{72}{336}$$

$$\Rightarrow \frac{336-72}{336} = \frac{264}{336} = 0.7857$$

$$\boxed{\therefore r_K = 0.7857}$$

∴ r>0, the rank correlation is said to be positive rank correlation.

When Ranks are not given:-

* Assign the ranks based on ascending (or) descending order.
* Assign the ranks to the following variables in the series
* Give the ranks

Apply formula, Rank correlation, $r_K = 1 - \frac{6\Sigma d^2}{N^3-N}$

1) Calculate the ~~Spparsman~~ Spearsman correlation from the following data.

years :	1	2	3	4	5	6	7
sales :	97.8	99.2	98.8	98.3	98.4	96.7	97.1
Prices :	73.2	85.8	78.9	75.8	77.2	87.2	83.8

Sol:- Assign the ranks based on descending i.e., highest to lowest.

Years	sales (x)	Rx	Prices (y)	Ry	d = (Rx-Ry)	d^2
1	97.8	5	73.2	7	-2	4
2	99.2	1	85.8	2	-1	1
3	98.8	2	78.9	4	-2	4
4	98.3	4	75.8	6	-2	4
5	98.4	3	77.2	5	-2	4
6	96.7	7	87.2	1	6	36
7	97.1	6	83.8	3	3	9
						$\Sigma d^2 = 62$

∴ N = 7

Rank correlation, $r_K = 1 - \dfrac{6\Sigma d^2}{N^3 - N}$

$$= 1 - \frac{6(62)}{7^3-7} = 1 - \frac{372}{343-7} = 1 - \frac{372}{336}$$

$$= 1 - 1.107$$

$$\boxed{\therefore r_K = -0.107.}$$

Conclusion: Here, r<0, then the co-efficient of rank correlation is negative rank correlation & the variables are negatively rank correlated.

When ranks are Equal:

Rank correlation, $r_K = 1 - \dfrac{6\left(\Sigma d^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + ----+\frac{1}{12}(m^3-m)\right]}{N^3 - N}$

where, m = no. of items rank is repeating in both the series

d = deviation between the ranks.

1) Explain the rank correlation co-efficient between the variables x and y from the following pairs of observed values.

x :	50	55	65	50	55	60	50	65	70	75
y :	110	110	115	125	140	115	130	120	115	160.

Sol: Here ranks are not given. We will assign the ranks based Ascending Order.

x	R_x	y	R_y	$d = R_x - R_y$	d^2
50	2	110	1.5	0.5	0.25
55	4.5	110	1.5	3	9
65	7.5	115	4	3.5	12.25
50	2	125	7	-5	25
55	4.5	140	9	-4.5	20.25
60	6	115	4	2	4
50	2	130	8	-6	36
65	7.5	120	6	1.5	2.25
70	9	115	4	5	25
75	10	160	10	0	0
					$\Sigma d^2 = 134.$

50 repeating '3' times = $\frac{1+2+3}{3} = \frac{6}{3} = 2$ gives equal rank to all the places where 50 is present.

55 repeating '2' times = $\frac{4+5}{2} = \frac{9}{2} = 4.5$ gives equal rank to all the places where 55 is present.

65 repeating '2' times = $\frac{7+8}{2} = \frac{15}{2} = 7.5$.

110 repeating '2' times = $\frac{1+2}{2} = \frac{3}{2} = 1.5$

115 repeating '3' times = $\frac{3+4+5}{3} = \frac{12}{3} = 4$.

m = no. of repeating ranks

= 3, 2, 2, 2, 3.

∴ Rank correlation,
$$r_k = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right]}{N^3-N}.$$

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

$$= 1 - \frac{6\left[134 + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3)\right]}{(10)^3-10}$$

$$= 1 - \frac{6\left[134 + \frac{1}{12}(27-3) + \frac{1}{12}(8-2) + \frac{1}{12}(8-2) + \frac{1}{12}(8-2) + \frac{1}{12}(27-3)\right]}{1000-10}$$

$$= 1 - \frac{6\left[134 + \frac{1}{12}\times 24 + \frac{1}{12}\times 6 + \frac{1}{12}\times 6 + \frac{1}{12}\times 6 + \frac{1}{12}\times 24\right]}{990}$$

$$= 1 - \frac{6[134+2+0.5+0.5+0.5\times 2]}{990} = 1 - \frac{6[134+5.5]}{990}$$

$$= 1 - \frac{6[139.5]}{990}$$

$$= 1 - \frac{837}{990}$$

$$= 1 - 0.844 \qquad 1 - 0.845$$

$$\boxed{\therefore r_K = 0.155}$$

Conclusion:- Here $r > 0$, then the co-efficient of rank correlation is positive rank correlation & the variables are positively rank correlated.

Concurrent Deviation Method:-

This method is very useful and very simplest method to calculate the correlation. In this method, to identity the direction of change of 'x' variable & 'y' variable.

Apply formula, $r = \pm\sqrt{\frac{\pm 2C - N}{N}}$

where, c = concurrent deviations. i.e., d = dx-dy +ve signs.

N = No. of pairs.

1) Calculate the co-efficient of concurrent deviation from the following data.

x:	60	55	50	56	30	70	40	35	80	80	75
y:	65	40	35	75	63	80	35	20	80	60	60.

Soln:

x	Dx	y	Dy	d=(dx×dy)
60		65		
55	−	40	−	+
50	−	35	−	+
56	+	75	+	+
30	−	63	−	+
70	+	80	+	+
40	−	35	−	+
35	−	20	−	+
80	+	80	+	+
80	0	60	−	0
75	−	60	0	0
				c = 8

∴ No. of positive signs, c=8; N=10 (no. of pairs is 10)

$$r = \pm\sqrt{\pm\frac{2c-N}{N}}$$

$$= \pm\sqrt{\pm\frac{2(8)-10}{10}} = \pm\sqrt{\pm\frac{16-10}{10}} = \pm\sqrt{\frac{6}{10}} = \pm\sqrt{\frac{3}{5}}$$

$$= \pm\sqrt{0.6}$$

$$\boxed{r = \pm 0.7745.}$$

Subject : Date :

Title of the test case :

Case study No. : Page No. :

Case study:4

1) 10 competitors in a beauty contest are ranked by 3 judges in the following order.

Judge-1 :	1	6	5	10	3	2	4	9	7	8
Judge-2 :	3	5	8	4	7	10	2	1	6	9
Judge-3 :	6	4	9	8	1	2	3	10	5	7

Use the rank correlation co-efficient to deteemine which pair of judges has the nearest approach to common tastes in beauty?

Sol:- In order to findout which pair of judges has the nearest approach to common tastes in beauty. we compare rank correlation between ~~first~~ judge & second judge & third judge.

i) calculate rank correlation between 1st & 2nd judges.

ii) calculate rank correlation between 2nd & 3rd judges.

iii) calculate rank correlation between 3rd & 1st judges

1st judge, 2nd judge & 3rd judge are considered as

R_1, R_2, R_3.

R_1	R_2	R_3	$D_1 = (R_1 - R_2)$	D_1^2	$D_2 (R_2 - R_3)$	D_2^2	$D_3 = (R_3 - R_1)$	D_3^2
1	3	6	-2	4	-3	9	5	25
6	5	4	1	1	1	1	-2	4
5	8	9	-3	9	-1	1	4	16
10	4	8	6	36	-4	16	-2	4
3	7	1	-4	16	6	36	-2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	-1	1
9	1	10	8	64	-9	81	1	1
7	6	5	1	1	1	1	-2	4
8	9	7	-1	1	2	4	-1	1
			~~$\Sigma D_1^2 = 200$~~	$\Sigma D_1^2 =$ 200	~~$\Sigma D_2^2 = 214$~~	$\Sigma D_2^2 = 214$		$\Sigma D_3^2 = 60$

$$r_k = 1 - \frac{6 \Sigma d^2}{N^3 - N}$$

$N = 10.$

(i) Rank correlation between 1st & 2nd judges

$$r_{12} = 1 - \frac{6 \Sigma d_1^2}{N^3 - N}$$

$$= 1 - \frac{6(200)}{10^3 - 10} = 1 - \frac{1200}{1000 - 10} = 1 - \frac{1200}{990} = 1 - 1.212$$

$$\boxed{\therefore r_{12} = -0.212}$$

(ii) Rank correlation between 2nd & 3rd judges

$$r_{23} = 1 - \frac{6 \Sigma d_2^2}{N^3 - N} = \frac{1 - 6(214)}{1000 - 10} = 1 - \frac{1284}{990}$$

Subject :
Date :
Title of the test case :
Case study No. :
Page No. :

$= 1 - 1.296$

$$\boxed{\therefore r_{23} = -0.296}$$

(iii) Rank correlation between 3rd & 1st judges

$$r_{31} = 1 - \frac{6\Sigma d_3^2}{N^3 - N} = 1 - \frac{6(60)}{10^3 - 10} = 1 - \frac{360}{1000-10} = 1 - \frac{360}{990}$$

$$= 1 - 0.3636$$

$$\boxed{\therefore r_{31} = 0.6364.}$$

Conclusion:- Using the rank correlation coefficient to 1st & 3rd judge has the nearest approach on common taste.

Regression:-

The word regression is used by the "Sir Frances Gatton" in the year 1877. Regression Analysis is the attempt to establish the relationship between two variable. In regression analysis, one variable is considered as dependent & other is independent.

Regression Equations:-

→ Regression equation x on y

$$x = a + by.$$

$$\Sigma x = Na + b\Sigma y.$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2.$$

Prepared By:
M.Navaneeth Kumar Reddy
Asst. Prof.

→ Regression equation y on x

$$y = a + bx$$

$$\Sigma y = Na + b\Sigma x.$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2.$$

1) From the following data obtain the two regression equation.

x : 6 2 10 4 8

y : 9 11 5 8 7

Sol:

x	y	x^2	y^2	xy
6	9	36	81	54
2	11	4	121	22
10	05	100	25	50
4	8	16	64	32
8	7	64	49	56
$\Sigma x = 30$	$\Sigma y = 40$	$\Sigma x^2 = 220$	$\Sigma y^2 = 340$	$\Sigma xy = 214$

Here, $N = 5$, $\Sigma x = 30$, $\Sigma y = 40$, $\Sigma x^2 = 220$, $\Sigma y^2 = 340$, $\Sigma xy = 214$.

Regression equation x on y:

$$x = a + by$$

$$\Sigma x = Na + b\Sigma y \rightarrow ①$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2 \rightarrow ②.$$

Substitute the values in the equations

$$30 = 5a + 40b \rightarrow ①$$

$$214 = 40a + 340b \rightarrow ②$$

Multiply the equation ① by ⑧.

Subject :		Date :	
Title of the test case :			
Case study No. :		Page No. :	

$① \times 8 \rightarrow 240 = 40a + 320b$

$214 = 40a + 340b$

$(-) \quad (-) \quad (-)$

$26 = -20b$

$-20b = 26$

$-b = \frac{26}{20}$

$-b = 1.3$

$\boxed{\therefore b = -1.3}$

Substitute $b = -1.3$ in the equation ①

$30 = 5a + 40(-1.3)$

$30 = 5a - 52$

$30 + 52 = 5a$

$82 = 5a \rightarrow a = \frac{82}{5} \rightarrow \boxed{a = 16.4.}$

Substitute a, b values in equation

$x = a + by$

$\boxed{\therefore x = 16.4 - 1.3y.}$

Regression equation Y on X :-

$y = a + bx$

$\Sigma y = Na + b\Sigma x \rightarrow ①$

$\Sigma xy = a\Sigma x + b\Sigma x^2 \rightarrow ②$

Substitute the values in equations ① & ②.

$40 = 5a + b(30) \rightarrow ①$

$214 = a(30) + b(220) \rightarrow ②.$

$40 = 5a + 30b \rightarrow ①$

$214 = 30a + 220b \rightarrow ②$

Multiply the equation ① by 6.

$$
\begin{aligned}
①\times 6 = 240 &= \cancel{30a} + 180b \\
214 &= \cancel{30a} + 220b \\
(-) \quad & \quad (-) \quad (-) \\
\hline
26 &= -40b
\end{aligned}
$$

$b = \frac{-26}{40}$

$\boxed{\therefore b = -0.65}$

Substitute 'b' value in equation ①

$40 = 5a + 30(-0.65)$

$40 = 5a - 19.5$

$5a = 40 + 19.5$

$5a = 59.5 \Rightarrow a = \frac{59.5}{5}$

$\boxed{\therefore a = 11.9}$

Substitute $a = 11.9,\ b = -0.65$ in equation

$y = a + bx$

$y = 11.9 - 0.65x$

$\boxed{\therefore y = 11.9 - 0.65x.}$

Subject :   Date :

Title of the test case :

Case study No. :   Page No. :

With the help of Mean:-

* Regression equation x on y $\Rightarrow x-\bar{x} = \frac{\Sigma xy}{\Sigma y^2}(y-\bar{y})$

* Regression equation y on x $\Rightarrow y-\bar{y} = \frac{\Sigma xy}{\Sigma x^2}(x-\bar{x})$.

1) From the following data obtain the gegression equations.

x : 6 2 10 4 8

y : 9 11 5 8 7.

Sol:-

x	y	$x=(x-\bar{x})$	$y=(y-\bar{y})$	x^2	y^2	xy
6	9	0	1	0	1	0
2	11	-4	3	16	9	-12
10	5	4	-3	16	9	-12
4	8	-2	0	4	0	0
8	7	2	-1	4	1	-2
$\Sigma x=30$	$\Sigma y=40$			$\Sigma x^2=40$	$\Sigma y^2=20$	$\Sigma xy=-26$.

Mean, $\bar{x} = \frac{\Sigma x}{N} = \frac{30}{5} = 6$.

$\boxed{\therefore \bar{x} = 6}$

· Mean, $\bar{y} = \frac{\Sigma y}{N} = \frac{40}{5} = 8$

$\boxed{\therefore \bar{y} = 8.}$

Regression equation x on y:-

x on y $\Rightarrow x-\bar{x} = \frac{\Sigma xy}{\Sigma y^2}(y-\bar{y})$

$$x-6 = \frac{-26}{20}(y-8)$$

$$x-6 = -1.3\ (y-8)$$

$$x-6 = -1.3y + 1.3(8)$$

$$x-6 = -1.3y + 10.4$$

$$x = -1.3y + 10.4 + 6$$

$$x = -1.3y + 16.4$$

$$\boxed{\therefore x = 16.4 - 1.3y.}$$

Regression equation y on x :-

$$y \text{ on } x \Rightarrow y-\bar{y} = \frac{\Sigma xy}{\Sigma x^2}(y-(x-\bar{x})$$

$$y-8 = \frac{-26}{40}(x-6)$$

$$y-8 = -0.65(x-6)$$

$$y-8 = -0.65x + 6(0.65)$$

$$y-8 = -0.65x + 3.90$$

$$y-8 = -0.65x + 3.9$$

$$y = -0.65x + 3.9 + 8$$

$$y = -0.65x + 11.9$$

$$y = 11.9 - 0.65x.$$

$$\boxed{\therefore y = 11.9 - 0.65x.}$$

Deviation | with the help of Assumed Mean :-

Regression equation x on y :-

$$(x-\bar{x}) = r.\frac{\bar{x}}{\bar{y}}(y-\bar{y})$$

$$r\frac{\bar{x}}{\bar{y}} = \frac{N\Sigma dxdy - \Sigma dx.\Sigma dy}{N\Sigma dy^2 - (\Sigma dy)^2}.$$

Subject : Date :

Title of the test case :

Case study No. : Page No. :

Regression equation y on x :-

$$y-\bar{y} = \bar{r}\cdot\frac{\bar{y}}{\bar{x}}(x-\bar{x})$$

$$r\cdot\frac{\bar{y}}{\bar{x}} = \frac{N\Sigma dx\,dy - \Sigma dx\cdot\Sigma dy}{N\Sigma dx^2 - (\Sigma dx)^2}.$$

1) obtain the regression equations from the following data.

x : 6 2 10 4 8.

y : 9 11 5 8 7.

Sol:- Assumed mean in x = 2, y = 5

N = 5.

x	y	$dx=(x-A)$	$dy=(y-A)$	dx^2	dy^2	$dx\,dy$.
6	9	4	4	16	16	16
2	11	0	6	0	36	0
10	5	8	0	64	0	0
4	8	2	3	4	9	6
8	7	6	2	36	4	12
$\Sigma x=30$	$\Sigma y=40$	$\Sigma dx=20$	$\Sigma dy=15$	$\Sigma dx^2=120$	$\Sigma dy^2=65$	$\Sigma dx\,dy=34$

Mean, $\bar{x} = \frac{\Sigma x}{N} = \frac{30}{5} = 6.$ Mean, $\bar{y} = \frac{\Sigma y}{N} = \frac{40}{5} = 8$

$\boxed{\therefore \bar{x} = 6}$ $\boxed{\therefore \bar{y} = 8.}$

Regression equation x on y :-

$$(x-\bar{x}) = r\cdot\frac{\bar{x}}{\bar{y}}(y-\bar{y}).$$

$$r\cdot\frac{\bar{x}}{\bar{y}} = \frac{N\Sigma dxdy - \Sigma dx\cdot\Sigma dy}{N\Sigma dy - (\Sigma dy)^2}$$

$$= \frac{5(34) - 20\times 15}{5(65) - (15)^2}$$

$$= \frac{170-300}{325-225} = \frac{-130}{100} = -1.3.$$

$$\boxed{\therefore r\cdot\frac{\bar{x}}{\bar{y}} = -1.3.}$$

Substitute $r\cdot\frac{\bar{x}}{\bar{y}}$ in equation x on y.

$$x - \bar{x} = r\cdot\frac{\bar{x}}{\bar{y}}(y-\bar{y})$$

$$x-6 = -1.3(y-8)$$

$$x-6 = -1.3y + 1.3(8)$$

$$x-6 = -1.3y + 10.4$$

$$x = -1.3y + 10.4 + 6$$

$$x = -1.3y + 16.4$$

$$\boxed{\therefore x = 16.4 - 1.3y.}$$

Regression equation y on x :-

$$(y-\bar{y}) = r\cdot\frac{\bar{y}}{\bar{x}}(x-\bar{x})$$

$$r\cdot\frac{\bar{y}}{\bar{x}} = \frac{N\Sigma dxdy - \Sigma dx\,\Sigma dy}{N\Sigma dx^2 - (\Sigma dx)^2}$$

$$= \frac{5(34) - 20(15)}{5(120) - (20)^2} = \frac{170-300}{600-400} = \frac{-130}{200}.$$

$$\boxed{\therefore r\cdot\frac{\bar{y}}{\bar{x}} = -0.65.}$$

Subject :
Title of the test case :
Case study No. :

Date :
Page No. :

Substitute $r \cdot \frac{\bar{y}}{\bar{x}}$ in regression equation y on x.

$$y-\bar{y} = \bar{r} \cdot \frac{\bar{y}}{\bar{x}} (x-\bar{x})$$

$$y-8 = -0.65 (x-6)$$

$$y-8 = -0.65x + 6(0.65)$$

$$y-8 = -0.65x+3.9$$

$$y = -0.65x+3.9+8$$

$$y = -0.65x+11.9$$

$$\boxed{\therefore y = 11.9 - 0.65x.}$$

Standard Error calculation with help of Regression Equations :-

- * calculate the regression equations with the help of deviation with Assumed mean.
- * Then calculate the standard error value.

$$x \text{ on } y \Rightarrow \sqrt{\frac{\Sigma(x-x_c)^2}{N}}$$

$$y \text{ on } x \Rightarrow \sqrt{\frac{\Sigma(y-y_c)^2}{N}}$$

x_c, y_c = change values (or) standard error value of the given variables.

1) Calculate standard error with the help of regression.

x : 6 2 10 4 8

y : 9 11 5 8 7.

Solution: Assumed mean in $x=2, y=5$.

x	y	$dx=(x-A)$	$dy=(y-A)$	dx^2	dy^2	$dxdy$.
6	9	4	4	16	16	16
02	11	0	6	~~36~~ 0	36	0
10	5	8	0	64	0	0
4	8	2	3	4	9	6
8	7	6	2	36	4	12
$\Sigma x=30$	$\Sigma y=40$	$\Sigma dx=20$	$\Sigma dy=15$	$\Sigma dx^2=120$	$\Sigma dy^2=65$	$\Sigma dxdy=34$

Mean, $\bar{x}=\frac{\Sigma x}{N}$

$=\frac{30}{5}$

$\boxed{\bar{x}=6.}$

Mean, $\bar{y}=\frac{\Sigma y}{N}$

$=\frac{40}{5}$

$\boxed{\bar{y}=8}$

Regression X on y :-

$$x-\bar{x}=r.\frac{\bar{x}}{\bar{y}}(y-\bar{y})$$

$$r.\frac{\bar{x}}{\bar{y}}=\frac{N\Sigma dxdy-\Sigma dx\Sigma dy}{N\Sigma dy^2-(\Sigma dy)^2}$$

$$=\frac{5(34)-20\times 15}{5\times 65-(15)^2}$$

$$=\frac{170-300}{325-225}$$

$$=\frac{-130}{100}=-1.3$$

$$\boxed{r.\frac{\bar{x}}{\bar{y}}=-1.3.}$$

Subject : Date :

Title of the test case :

Case study No. : Page No. :

Substitute $r.\frac{\overline{x}}{\overline{y}}$ in equation x on y.

$$x-6 = -1.3(y-8)$$

$$x-6 = -1.3y + 1.3(8)$$

$$x-6 = -1.3y + 10.4$$

$$x = -1.3y + 10.4 + 6$$

$$\boxed{\therefore x = 16.4 - 1.3y}$$

Regression Y on X :-

$$y-\overline{y} = r.\frac{\overline{y}}{\overline{x}}(x-\overline{x})$$

$$r.\frac{\overline{y}}{\overline{x}} = \frac{N\Sigma dxdy - \Sigma dx.\Sigma dy}{N\Sigma dx^2 - (\Sigma dx)^2}$$

$$= \frac{5(34) - 20 \times 15}{5(120) - (20)^2} = \frac{170 - 300}{600 - 400} = -\frac{130}{200} = -0.65$$

$$\boxed{r.\frac{\overline{y}}{\overline{x}} = -0.65}$$

Substitute $r.\frac{\overline{y}}{\overline{x}}$ value in Y on X.

$$y-8 = r.\frac{\overline{y}}{\overline{x}}(x-6)$$

$$y-8 = -0.65(x-6)$$

$$y-8 = -0.65x + 6(0.65)$$

$$y-8 = -0.65x + 3.9.$$

$$y = -0.65x + 3.9 + 8$$

$$\boxed{\therefore y = 11.9 - 0.65x}.$$

x on y - standard errors :-

$x = 16.4 - 1.3y$ $\qquad$ $y = 9, 11, 5, 8, 7.$

Substitute y values in $x = 16.4 - 13 \cdot y.$

$9 \Rightarrow 16.4 - 1.3(9)$

$= 16.4 - 11.7.$

$x_c = 4.7.$

$11 \Rightarrow 16.4 - 1.3(11)$

$= 16.4 - 14.3$

$x_c = 2.1.$

$5 \Rightarrow 16.4 - 1.3(5)$

$= 16.4 - 6.5$

$x_c = 9.9$

$8 \Rightarrow 16.4 - 1.3(8)$

$= 16.4 - 10.4$

$x_c = 6$

$7 \Rightarrow 16.4 - 1.3(7)$

$= 16.4 - 9.1$

$x_c = 7.3$

$y = 11.9 - 0.65x$ $\qquad$ $x = 6, 2, 10, 4, 8.$

Substitute x values in $y = 11.9 - 0.65x.$

$6 \Rightarrow 11.9 - 0.65(6)$

$= 11.9 - 3.90$

$y_c = 8$

$2 \Rightarrow 11.9 - 0.65(2)$

$= 11.9 - 1.3$

$y_c = 10.6$

Subject : Date :

Title of the test case :

Case study No. : Page No. :

$10 \Rightarrow 11.9 - 0.65(10)$

$= 11.9 - 6.5$

$y_c = 5.4.$

$4 \Rightarrow 11.9 - 0.65(4)$

$= 11.9 - 2.6$

$y_c = 9.3$

$8 \Rightarrow 11.9 - 0.65(8)$

$= 11.9 - 5.20$

$y_c = 6.7.$

x	y	x_c	y_c	$(x-x_c)$	$(y-y_c)$	$(x-x_c)^2$	$(y-y_c)^2$
6	9	4.7	8	1.3	1	1.69	1.
2	11	2.1	10.6	-0.1	0.4	0.01	0.16
10	5	9.9	5.4	0.1	-0.4	0.01	0.16
4	8	6	9.3	-2	1.3	4	1.69
8	7	7.3	6.7	0.7	0.3	0.49	0.09
						$\Sigma(x-x_c)^2 =$ 6.20	$\Sigma(y-y_c)^2 =$ 3.10.

$\Sigma(x-x_c)^2 = 6.2$

$\Sigma(y-y_c)^2 = 3.1.$

$$x \text{ on } y = \sqrt{\frac{\Sigma(x-x_c)^2}{n}} = \sqrt{\frac{6.2}{5}} = \sqrt{1.24} = 1.113.$$

$$y \text{ on } x = \sqrt{\frac{\Sigma(y-y_c)^2}{n}} = \sqrt{\frac{3.1}{5}} = \sqrt{0.62} = 0.737.$$

Standard error in x on y = 1.113.

Standard error in y on x = 0.737.

BIMK

Subject :
Title of the test case :
Case study No. :
Date :
Page No. :

UNIT - 3

PROBABILITY

Meaning and Definition of Probability :-

* The word probability is very commonly used in day-to-day conversation and generally people have no clear idea about its meaning.

* The probability of a given event is a "expression of chance of occurance of an event."

* Probability is a number which ranges from '0' to '1'.
'0' is a failure case (or) event not occur.
'1' is a success case (or) event can occur.

According to American heritage dictionary "Probability is the branch of mathematics that studies the chance of occurance of random events in order to predict the behaviour of a defined system".

Significance of Probability in Business Applications :-

* Probability theory has been developed & employed to treat and solve many weighting problems

* Probability is the foundation of the classical decision procedures of estimation & testing.

* Probability models can be very useful for making predictions
* Probability is concerned with the construction of econometric models with managerial decisions on planning and control with the occurance of accidents, of all kinds & with random disturbances in an electrical mechanism.
* Probability is involved in the observation of the life span of a radio active atom.

→ The phenotypes of the offspring.

→ The crossing of two species of plants.

→ The discussion about sex of an unborn baby etc.,

* Probability has become an indispensable tool for all types of formal studies that involve uncertainity.
* It should be noted that the concept of probability is employed not only for various types of scientific investigations, but also for many problems in everyday life.
* The probability theory provides a media of coping up with uncertainity.
* Highlighting the importane of probability theory is a method of decisions making under uncertainity.

Note :- Formula for getting the Probability.

$$P(E) = \frac{\text{Number of favourable cases}}{\text{Total number of likely cases}}$$

$$P(E) = \frac{P(S)}{P(N)}.$$

where, $p(s)$ = favouqable cases i.e., $^{n}C_{r}$

$p(N)$ = Total no. of cases i.e., $^{N}C_{r}$.

1) A bag contains 10 black & 20 white balls, a ball is drawn at random. what is the probability that it is black?

Sol: Total no. of balls in a bag = 20 white + 10 black balls
= 30 balls.

No. of black balls = 10

No. of white balls = 20.

what is the probability of getting a black ball.

$$P(E) = \frac{10_{C_1}}{30_{C_1}}$$

$$= \frac{10}{30}$$

$$= 1/3$$

$$= 0.333$$

$$\boxed{P(E) = \frac{1}{3} \text{(or)} 0.333.}$$

what is the probability of not getting a black ball i.e.,

$= 1 - P(E)$

$= 1 - 0.33$

$= 0.67.$

Note:- Sum of the probability is equal to '1'.
i.e., combination of both success and failure case.

$$p + q = 1.$$

where, p= success case

q =failure case.

* The probability getting a success case is 'p' is known. we can get the failure case

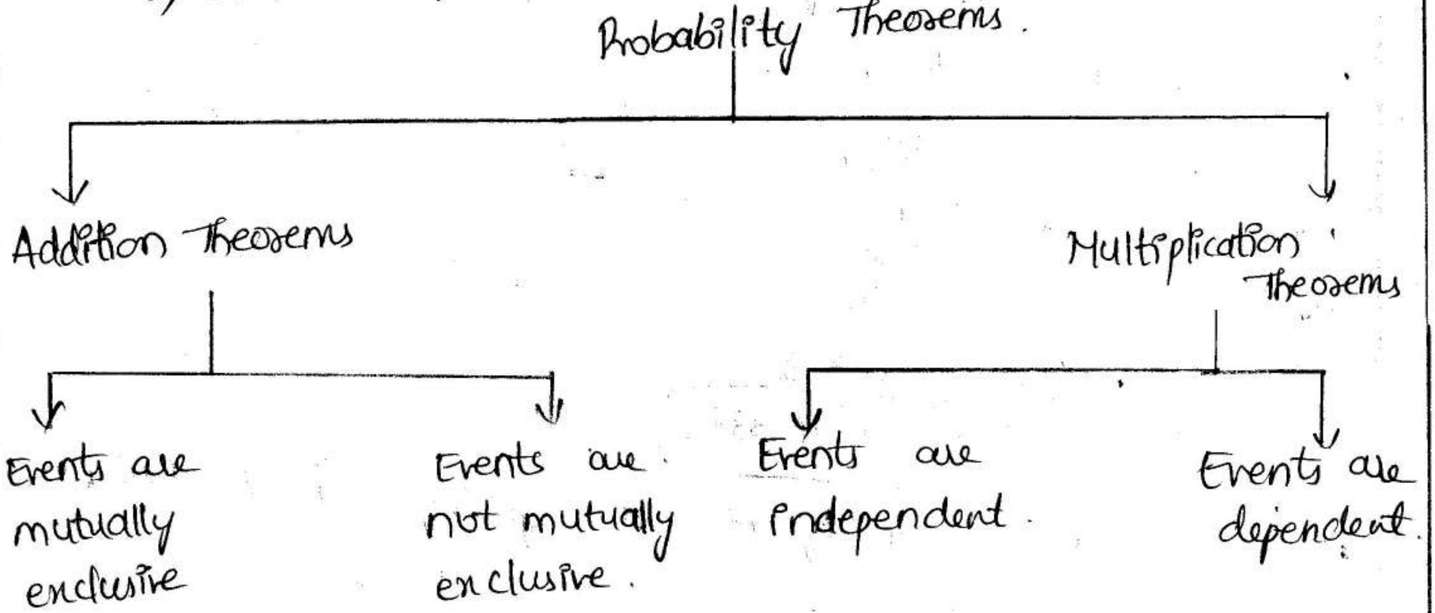
$$q = 1-p.$$

i.e., failure case is equal to difference between the sum of the probabilities to the success case.

Theories of Probability :-

These are 2 types of theories of probability namely.

1) The addition theorem.
2) The multiplication theorem.



Addition Theorems :-

This rule is related to the addition operation between two types of events to occur.

Subject	:	Date	:
Title of the test case	:		
Case study No.	:	Page No.	:

1) Mutually Exclusive Events :-

A and B are mutually exclusive the probability of the occurance of either A (or) B is the sum of individual probability of A & B.

Symbolically: $P(A \text{ or } B) = P(A) + P(B)$

(or)

$P(A \cup B) = P(A) + P(B)$

Proof of the Theorem :-

If an event A can happen in a_1 ways & B in a_2 ways, then the number of ways in which either event can happen is $a_1 + a_2$. If the total no. of probabilities is n, then by definition the probability of either the first (or) the second event happening is

$$\frac{a_1 + a_2}{n} = \frac{a_1}{n} + \frac{a_2}{n}$$

But, $\frac{a_1}{n} = P(A)$ & $\frac{a_2}{n} = P(B)$

Here, $P(A \text{ or } B) = P(A) + P(B)$ the theorem expand.

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C).$$

2) When events are not mutually Exclusive :-

when events are not mutually exclusive (or) in other words, it is possible for both events to occur, the addition rule must be modified.

Here, for finding the probability of one (or) more of two events that are not mutually exclusive we use the modified form of the addition theorem.

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$p(A \cup B)$ = probability of A or B happening when A or B are not mutually exclusive.

$$p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(A \cap B) - p(B \cap C) - p(A \cap C) + p(A \cap B \cap C)$$

Mutually Exclusive Events: sums.

1) One card is drawn from a standard pack of 52. What is the probability that it is either a king or a queen?

Sol: There are 4 kings, 4 queens in a pack of 52 cards.

The probability that the card is drawn as a king;

i.e., $\frac{4c_1}{52c_1} = \frac{4}{52} = \frac{1}{13}$.

$P(A) = \frac{1}{13}$.

The probability that the card is drawn as a queen,

i.e., $\frac{4c_1}{52c_1} = \frac{4}{52} = \frac{1}{13}$

$P(B) = \frac{1}{13}$.

Since, the events are mutually exclusive, the probability that the card drawn is either a king (or) queen.

i.e., $P(A \cup B) = p(A) + P(B)$

$= \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$

Subject :

Title of the test case :

Case study No. :

Date :

Page No. :

$\boxed{\therefore P(A \cup B) = 0.1538}$

Events are mutually not Exclusive :-

1) The managing committee of vishali welfare association formed a sub-committee of 5 persons to look into electricity problem. Profiles of 5 persons are mutually not exclusive.

1) Male age 40
2) Female age 27.
3) Male age 43.
4) Male age 65.
5) Female age 38.

If a chair person has to be selected from this what is the probability that he ~~could be~~ either female or over the 30 years?

Sol:- $p(\text{female or over } 30) = p(\text{female}) + p(\text{over } 30) - p(\text{female \& over } 30)$

probability of female $p(\text{female}) = \frac{2c_1}{5c_1} = \frac{2}{5}$.

Probability of ~~fe~~ over 30, $p(\text{over } 30) = \frac{4c_1}{5c_1} = \frac{4}{5}$

Probability of female & over 30, $p(\text{female and over } 30) = \frac{1c_1}{5c_1} = \frac{1}{5}$

$$\therefore p(\text{female or over } 30) = p(\text{female}) + p(\text{over } 30) - p(\text{female and over } 30)$$

$$= \frac{2}{5} + \frac{4}{5} - \frac{1}{5}$$

$$= \frac{6}{5} - \frac{1}{5} = \frac{5}{5} = 1.$$

Multiplication Theorem :- (Events are Independent).

This theorem states that if two events A & B are independent, the probability that they both will occcq is equal to the product of their individual probability.

Symbolically, if A and B are independent, then

$$p(A\cap B) = p(A) \times p(B)$$

$$p(A\cap B\cap C) = p(A) \times p(B) \times p(C).$$

Proof of the Theorem :-

If an event A can happen in n_1 ways of which a_1 are successful and the events B can happen in n_2 ways of which a_2 are successful. we can combine each successful event in the first with each successful event in the second case. Thus, the total no. of successful happenings in both cases is $a_1 \times a_2$. Similarly, the total no. of possible cases is $n_1 \times n_2$.

Then by definition the probability of the occurance of both events is

$$\frac{a_1 \times a_2}{n_1 \times n_2} = \frac{a_1}{n_1} \times \frac{a_2}{n_2}$$

we know $\frac{a_1}{n_1} = p(A)$; $\frac{a_2}{n_2} = p(B)$

$$p(A\cap B) = p(A) \times p(B)$$

Events are Independent :-

1) A man wants to marry a girl having qualities.

(i) white complextion - The probability of getting such a girl is one in twenty.

Subject :

Title of the test case :

Case study No. :

Date :

Page No. :

(ii) Handsome dowry - The probability of getting such a person is 1 in 50.

(iii) Westernized manner - The probability of getting such a person is 1 in 100.

Find out the probability of his getting married to such a girl when the person of these 3 attributes is Independent

Sol:- The probability of getting a girl with white complextion, $P(A) = \frac{1}{20}$

The probability of getting a girl with handsome dowry, $P(B) = 1/50$

The probability of getting a girl with westernized manner, $P(C) = \frac{1}{100}$

The probability of getting all qualities held in one person simultaneously i.e., $P(A \cap B \cap C)$

$$\therefore P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

$$= \frac{1}{20} \cdot \frac{1}{50} \; \frac{1}{100}$$

$$= \frac{1}{1000 \times 100}$$

$$= \frac{1}{1,00,000}$$

$$= 0.00001$$

$$\boxed{\therefore P(A \cap B \cap C) = 0.00001}$$

Conditional Probability :- (Events are dependent).

The multiplication theorem explained above is not applicable in case of dependent events. Two events A & B are said to be dependent when B can occur only when A is known to have occured. the probability attached to such an event is called the "conditional Probability" & is denoted by "p(A|B)."

If two events A and B are dependent, then the conditional probability of B given A is

$$p(B|A) = \frac{p(A \cap B)}{p(A)} \Rightarrow p(A \cap B) = p(A) \times p(B|A)$$

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \Rightarrow p(A \cap B) = p(B) \times p(A|B).$$

1) A bag contains 5 white and 3 black balls. 2 balls are drawn at random one after the another without replacement. Find the probability that both balls drawn are black.

Sol:- The probability of drawing a black ball in the first attempt is $p(A) = \frac{3c_1}{8c_1}$

$$= \frac{3}{8}.$$

The probability of drawing the second ball is black. Given that the first ball is drawn black, $P(B|A) = \frac{2c_1}{7c_1} = \frac{2}{7}.$

$\therefore$ the probability that the both balls drawn the black is given by $p(A \cap B) = P(A) \cdot P(B|A)$

$$= 3/8 \times 2/7$$

Subject :　　　　　　　　Date :

Title of the test case :

Case study No. :　　　　　　　　Page No. :

$= \frac{6}{56}$

$= 0.1071.$

$\boxed{\therefore P(A \cap B) = 0.1071.}$

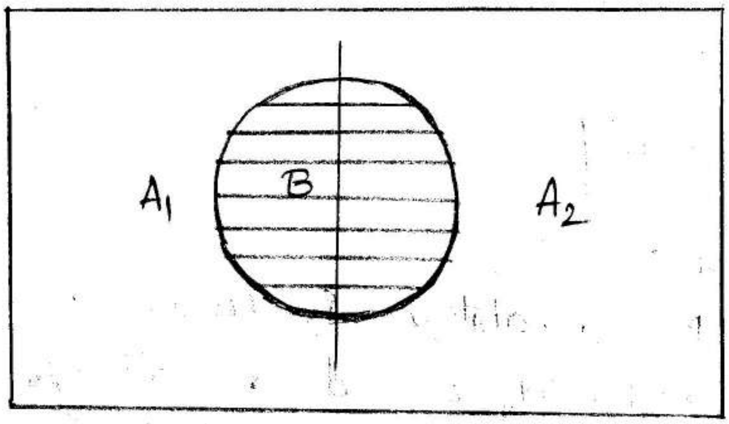
Baye's Theorem :-

The probability is known in different names, posterior probability, revised probability & inverse probability. This has been introduced by "Thomas Bayes" an english mathematician in this work known as Bayes in Decesion theory published in 1763. This theory consists of finding the probability of an event taking into account of a given sample information.

Baye's theorem is a means for qualifying uncertainity. Based on the probability theory, the theorem defines a rule for refining a hypothesis by factoring in additional evidence and back ground information and leads to a number representing the degree of probability that the hypothesis is true.

Thus a sample of 3 defective items out of 100 might be used to estimate the probability that a machine is (event A) not working properly (event B.)

It is to be denoted that the Bayesian probability is based on the formula of conditional probability where A_1 & A_2 are two events which are mutually exclusive & exhaustive & B is a simple event which intersects each of the A events as shown in the venn diagram to the right.



This is called "Posterior Probability" because it is calculated after information is taken into account. This is called "Revised Probability" as it is determined by revising the prior probabilities in the light of the additional information gathered. Further, this is called "Inverse Probability" also, as it consists of finding the probability of a problem.

However, the Bayesion (or) the posterior probabilities are always conditional probabilities which are calculated for every events as follows.

Mutually Exclusive Events :-

If an event E can only occur in combination with one of the mutually exclusive events $E_1, E_2, --- E_n$ then

$$p(E_k) = \frac{[p(E_k)][p(E/E_k)]}{\sum_{i=1}^{n} p(E_1)\, p(E/E_1)} \text{ ; where } k = 1, 2, ---- n$$

Mutually Exclusive & Exhaustive Events :-

If A_1, A_2 are two mutually exclusive and exhaustive events.

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(A_1) \cdot P(B/A_1) + P(A_2)\, P(B/A_2)}$$

$$P(A_2|B) = \frac{P(A_2)\, P(B/A_2)}{P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2)}$$

1) Assume that a factory has 2 machines past records. shows that machine 1 produces 30% of the items of the output and machine 2 produces 70% of the items from the output further 5% of items produced by machine 1 were defective & only 1% produced by machine 2 were defectives. If a defective item is drawn at random. what is the probability that the defective items produced by machine 1 (or) machine 2.

Sol:- Let A_1 = items produced by machine 1.

A_2 = items produced by machine 2.

B = defective items produced by either 1 (or) 2 machines.

Probability of the items produced by machine 1

$$P(A_1) = 30\% = \frac{30}{100} = 0.3.$$

Probability of the items produced by machine 2

$$P(A_2) = 70\% = \frac{70}{100} = 0.7.$$

The probability of the defective items in machine 1

$$P(B|A_1) = 5\% = \frac{5}{100} = 0.05$$

The probability of the defective items in machine 2

$$P(B|A_2) = 1\% = \frac{1}{100} = 0.01.$$

Probability of the defective items produced by machine 1

$$P(A1/B) = \frac{P(A_1)\cdot P(B|A_1)}{P(A_1)\cdot P(B|A_1) + P(A_2)\cdot P(B|A_2)}$$

$$= \frac{0.3 \times 0.05}{0.3 \times 0.05 + 0.7 \times 0.01}$$

$$= \frac{0.015}{0,015 + 0.007}$$

$$= \frac{0.015}{0.022}$$

$$P(A_1|B) = 0.68$$

Probability of defective items produced by machine 2

$$P(A_2|B) = \frac{P(A_2)\cdot P(B|A_2)}{P(A_1)\cdot P(B|A_1) + P(A_2)\cdot P(B|A_2)}$$

$$= \frac{0.7 \times 0.01}{0.3 \times 0.05 + 0.7 \times 0.01}$$

$$= \frac{0.007}{0.015 + 0.07}$$

$$= \frac{0.007}{0.022}$$

$$P(A_2|B) = 0.32$$

Subject : Date :
Title of the test case :
Case study No. : Page No. :

2) In a bolt factory machine A_1, machine A_2 and machine A_3 manufactures respectively 25%, 35% & 40% of the total of their output 5,4,2 percentages are defective bolts produced by the machines. A bolt is drawn at a random from the product is found to defective. what is the probability that it was manufactured by machine 3?

Sol:

$P(A_1) = 25\% = \frac{25}{100} = 0.25$

$P(A_2) = 35\% = \frac{35}{100} = 0.35$

$P(A_3) = 40\% = \frac{40}{100} = 0.40$

The defective items produced by machine A_1 $P(B|A_1) = 5\%$

$= \frac{5}{100} = 0.05$

The defective items produced by machine A_2 $P(B|A_2) = 4\%$

$= \frac{4}{100} = 0.04$

The defective items produced by machine A_3 $P(B|A_3) = 2\%$

$= \frac{2}{100} = 0.02$

The probability of defective items by machine A_3 is

$$P(A_3|B) = \frac{P(A_3).\ P(B|A_3)}{P(A_1)\cdot P(B|A_1) + P(A_2)\cdot P(B|A_2) + P(A_3).\ P(B|A_3)}$$

$$= \frac{0.4 \times 0.02}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= \frac{0.08}{0.0125+0.014+0.08}$$

$$= \frac{0.008}{0.0345}$$

$P(A_3|B) = 0.231$

Needs of Bayes Theorem :-

* The sample space is partioned into a set of mutually exclusive events $(A_1, A_2, ---- A_n)$.

* With in the sample space, there exists an event B for which $P(B) > 0$.

* The analytical goal is to compute a conditional probability of the form $P(A_K|B)$.

* Atleast one of the two sets of probabilities descussed below :

(i) $P(A_K \cap B)$ for each A_K.

(ii) $P(A_K)$ and $P(B|A_K)$ for each A_K.

Features :-

* Through it deals with a conditional probability; its intepretation is different form that of the general conditional probability theorem.

* Very useful to decision making.

* The nations of priors and posterior in Bayes theorem are relative to a given sample a outcome.

Applications :-

* The theorem still prescribes multiplying the prior distribution by the likelihood function and them normalising,