

Tuesday
3/10/17

BUSINESS STATISTICS

Introduction:-

Curiosity is the distinctive feature of human beings. Curiosity to know about ourselves, our environment, our planets etc., exist in human beings. Whenever such questions arise, the human beings seek answers to them. Whenever we face Problems, we try to find solutions to them. Seeking answers to questions and finding solutions to Problems has been the basis for human Progress.

"A Systematic Search for an answer to a question or a solution to a Problem is called RESEARCH."

Research simply Means — Search for facts.

— Solutions to Problems.

— Answers to Questions

— Purposive Investigation

— Organised Enquiry.

Statistics and the utility of statistics has been continuously used today in various fields. Statistics is a Science of Statistical Methods, tools and Techniques used in collecting, Presenting, Interpreting of the "DATA" in Managerial Decision Making.

- * Statistics may be called as the Science of counting.
- * Statistics is a Method of Decision Making in the face of Uncertainty and Risk.
- * Statistics is a method of Decision Making during Uncertainty on the basis of Numerical Data.
- * For any statistical Enquiry whether it is business, Economics

or Social Sciences, the basic problem is to collect facts and figures related to study. The person who conducts the statistical enquiry is called INVESTIGATOR and the persons from whom the information is collected are called RESPONDENTS. The entire statistical analysis is based on systematic collection of data.

DATA :-

The search for answers to questions is called collection of data. Data are facts and other relevant material, past and present, serving as the base for any statistical analysis.

- Ex:- 1. The marks obtained by the 1st sem MBA students in the subject Business Statistics.
2. The opinion of people in voting in General Elections.

Importance of DATA:-

* The Data serve as the basis or raw material for Analysis.

Without an analysis of factual data, no inferences can be drawn on the questions under study. Inferences based on imagination cannot provide correct answers to the research questions. The Relevance, Adequacy, Reliability of data determines the quality of the findings of the study.

Sources of DATA:-

Sources of data may be classified into

1. Primary Sources.
2. Secondary Sources.

Primary Sources are the original sources from which the researcher directly collects the data that has not been previously collected.

-for Example:- collection of data by directly by the researcher on brand awareness.

The Secondary Sources are the Sources containing the data which have been collected and compiled for another Purpose. The Secondary Sources consists of readily available information in the form of reports, financial statements etc., which will be used by the Researcher for their Studies.

Types of DATA:-

1. Qualitative Data [categorical Data].

2. Quantitative Data

Wednesday
4/10/17

Qualitative Data:-

Qualitative Data includes labels or Names used to identify the attributes (features / characteristics) of each element. Qualitative Data is measured in Nominal or Ordinal Scale. A Qualitative variable is a variable with Qualitative Data which gives responses like 'YES' or 'NO'.

Ex:- Do you like soft drinks?

The Answer may be limited to YES or NO.

In Qualitative Data, the Statistical Analysis is limited i.e., we can summarize the Qualitative Data by counting the Number of Observations in each category. The Qualitative Data when grouped in different categories can be applied with Arithmetic operations like Addition, Subtraction, Multiplication and Division which will provide meaningful results.

Quantitative Data:-

Quantitative Data require Numerical values that indicate "How Much or How Many". Quantitative Data is obtained using the ^{grouping} interval or Scale of Measurement. Qualitative Data will provide scope for Arithmetic operations. For example:- for a Quantitative Data the data may be added, subtracted, multiplied and divided for meaningful Response. In general more alternatives for Quantitative Data.

Analysis using statistics is possible.

Qualitative Data	Quantitative Data
* Classification of population by Gender (Male / Female).	* The marks obtained by students
* Students Performance in the Examinations (Good / bad).	* The runs scored by a cricketer. (Rohit Sharma)
* Persons in rich and poor	* The number of votes obtained by a Political Party

Summarizing Qualitative Data:-

Consider the following example to demonstrate the construction and interpretation of a frequency Distribution for a Qualitative Data. Consider five soft drinks like Coke, Pepsi, Merinda, Maza and Sprite. Assume the data given below. Construct frequency distribution table and find out the relative frequency, Percent frequency, Draw Bar Graphs and Pie charts.

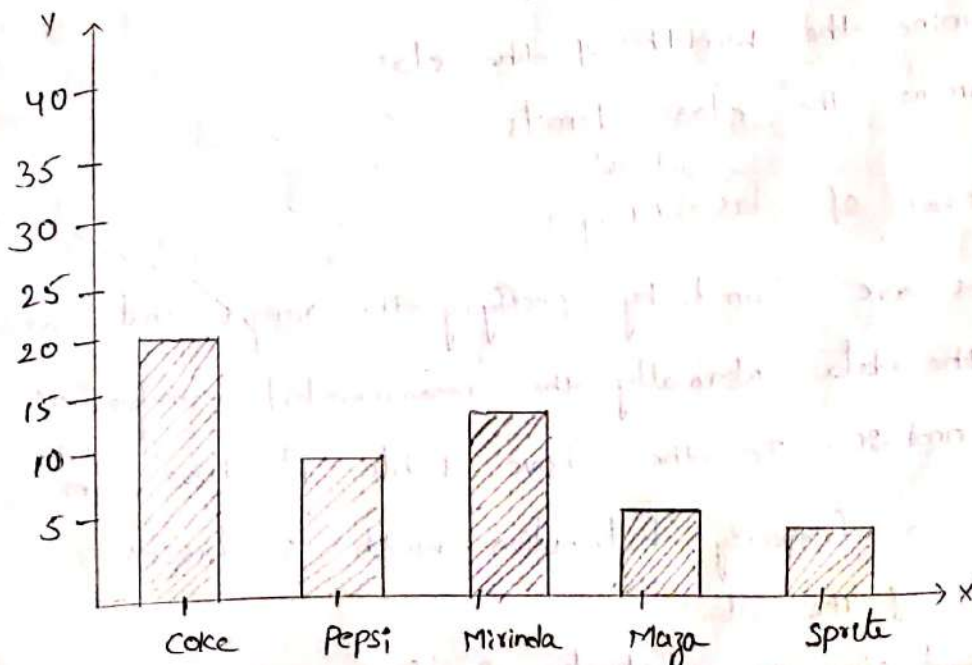
Coke	Pepsi	Merinda	Coke	Pepsi
Coke	Maza	Pepsi	Merinda	Merinda
Coke	Maza	Sprite	Coke	Pepsi
Coke	Coke	Sprite	Coke	Pepsi
Coke	Pepsi	Coke	Sprite	Merinda
Coke	Coke	Coke	Merinda	Coke
Sprite	Maza	Merinda	Pepsi	Merinda
Coke	Coke	Coke	Merinda	Maza
Coke	Pepsi	Merinda	Merinda	Merinda
Merinda	Coke	Maza	Merinda	Coke

Soft Drinks	Frequency
Coke	20
Pepsi	8
Mirinda	13
Maza	5
Sprite	4

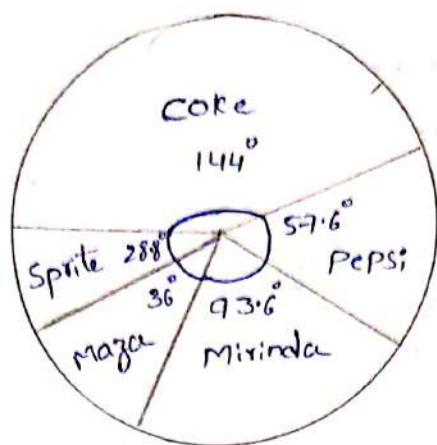
From the above table it is understood that Coke is the first Preference, Mirinda is the Second Preference, Pepsi the third Preference, Maza the fourth and Sprite the last.

Soft Drinks	Relative frequency	Percent frequency	Degrees
Coke	$\frac{20}{50} = 0.4$	40	$0.4 \times 360^\circ = 144^\circ$
Pepsi	$\frac{8}{50} = 0.16$	16	$0.16 \times 360^\circ = 57.6^\circ$
Mirinda	$\frac{13}{50} = 0.26$	26	$0.26 \times 360^\circ = 93.6^\circ$
Maza	$\frac{5}{50} = 0.1$	10	$0.1 \times 360^\circ = 36^\circ$
Sprite	$\frac{4}{50} = 0.08$	8	$0.08 \times 360^\circ = 28.8^\circ$

Bar Graph



Pie chart



Summarizing Quantitative Data:-

Year End Audit Time (in days)			
12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

Consider the above Quantitative Data given in the table. This table shows the time in days required to complete year end Audit for a sample of 20 companies. The following are necessary to be calculated to construct a frequency distribution table with the Quantitative Data

- i) Determine the number of classes.
- ii) Determine the width of the class.
- iii) Determine the class limits.

Friday
6/10/17 i) Number of classes (Range) $\xrightarrow{\text{Interval}}$

The classes are formed by specifying the ranges that will be used to group the data. Normally the recommended number of classes is between 5 and 20. In the above table we have $n = 20$, so we can choose a frequency distribution with 5 classes

- ii) Width of the class

The second step in constructing a frequency distribution for Quantitative

Data is choosing the width of the class. As a general Rule it is recommended the width of the classes be same. To determine the width of the class we should adopt the following formula

$$\text{Approximate class width} = \frac{\text{Largest data value} - \text{Smallest data value}}{n(\text{number of classes})}$$

$$= \frac{33-12}{5}$$

$$= \frac{21}{5} = 4.2 \approx 5.$$

iii) Class Limits:-

Class limits may be chosen so that each data item belong to only one class. The lower class limit identifies the smallest possible data and the upper class limit identifies the largest possible data.

Frequency Distribution

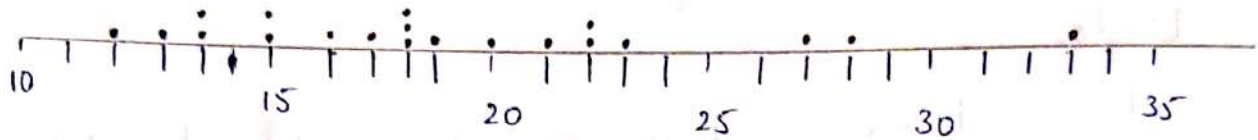
Addit time in days	frequency
10-14	4
15-19	8
20-24	5
25-29	2
30-34	1
	20

Relative frequency

Addit time	Relative frequency	Percent frequency
10-14	$4/20 = 0.2$	20
15-19	$8/20 = 0.4$	40
20-24	$5/20 = 0.25$	25
25-29	$2/20 = 0.1$	10
30-34	$1/20 = 0.05$	5
		100

Dot plot :-

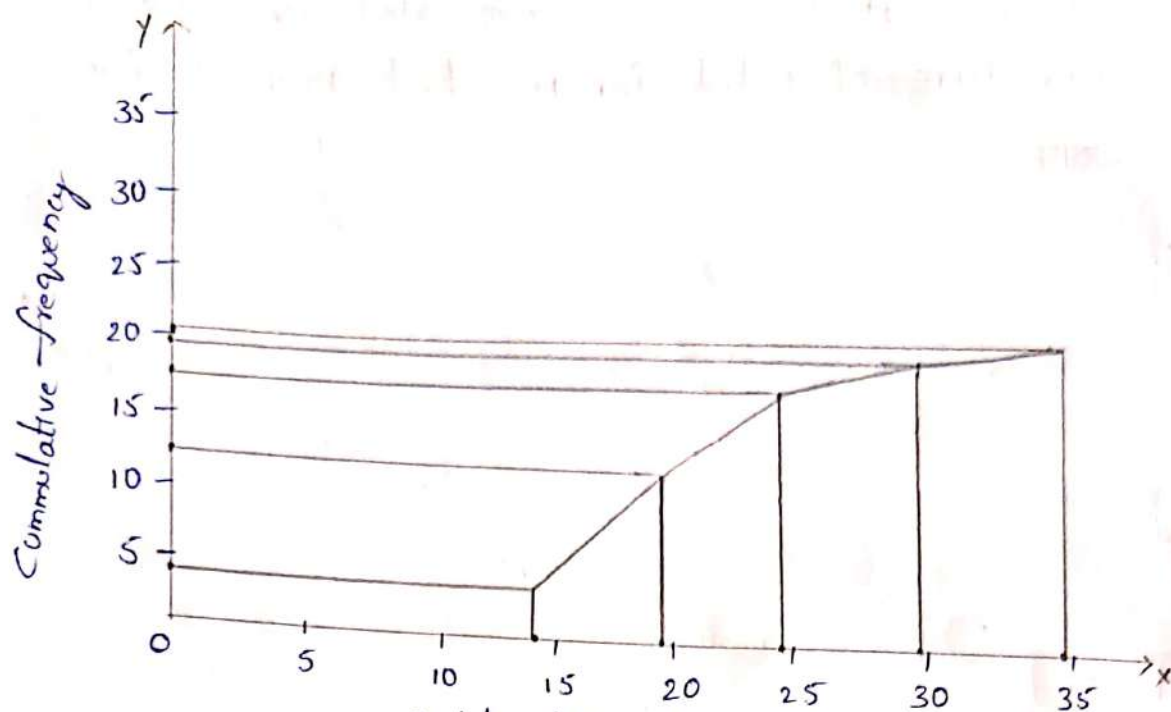
A dot plot is one of the simplest graphical summary of the data. The horizontal axis is drawn which shows the range of values for the observations. Each data item is represented by DOT placed above the axis.



OGIVE Curve:-

A Graph or a cumulative distribution is called OGIVE which shows the data values on x-axis and either the cumulative frequency or cumulative relative frequency or cumulative Percent frequency on y-axis.

Audit time in days	frequency	cumulative frequency	Relative frequency	cumulative Relative frequency	Percent frequency	cumulative Percent frequency
less than or upto 14	4	4	0.20	0.20	20	20
≤ 19	8	12	0.40	0.60	40	60
24	5	17	0.25	0.85	25	85
29	2	19	0.10	0.95	10	95
34	1	20	0.05	1.00	5	100



Cross Sectional And Time Series Data:-

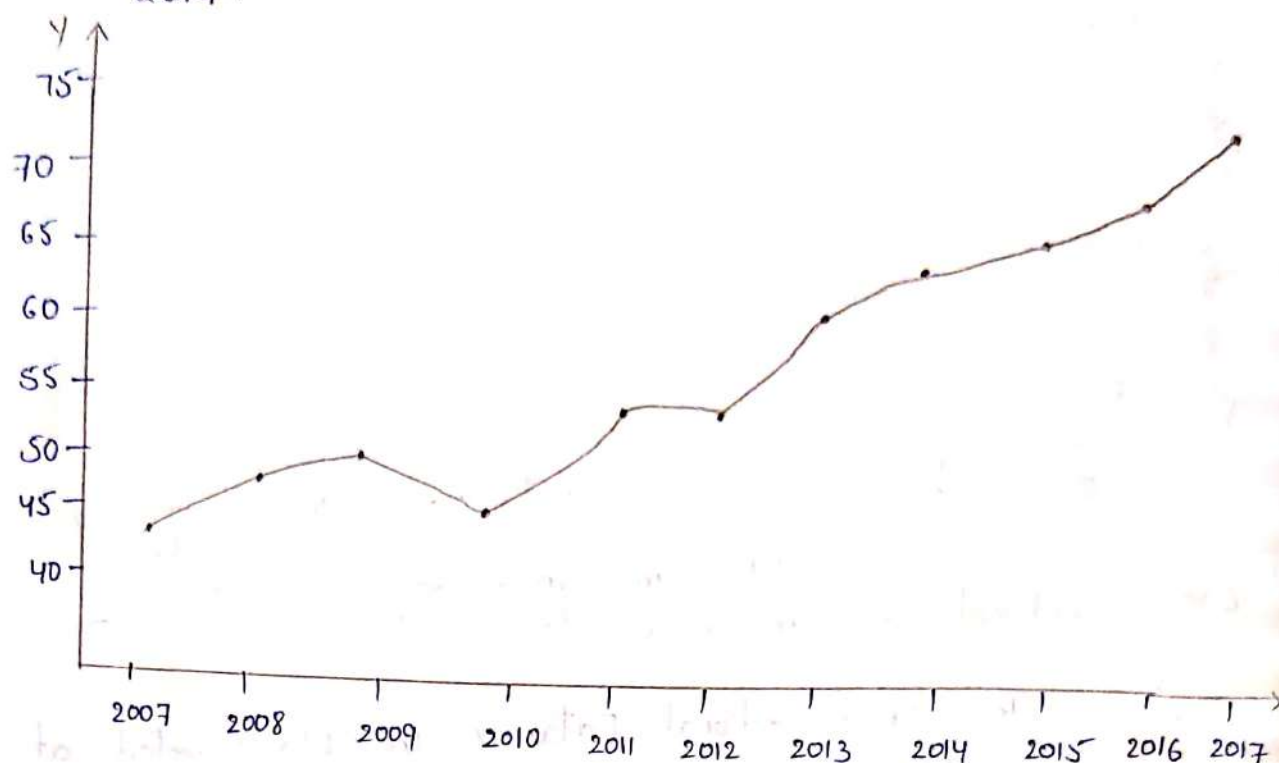
The cross Sectional Data is the data collected at the same or approximately same point of time.

Ex:- The Rainfall intensity on a specific date in 10 different places in a district (or) The number of candidates admitted into MBA in 10 different colleges affiliated to S.V. University in the year 2017.

College	MBA Admissions
A	180
B	230
C	162
D	115
E	84
F	60
G	114
H	172

Time Series data is the data collected over several time periods.

Ex:- The following Graph provides information about the average Price per litre of petrol for the last 10 years from 2007 to 2017.



Data Sources:-

The data can be obtained from

1. Existing Sources
2. Statistical Sources / Studies
 - Observational
 - Experimental.

1. Existing Sources:-

In Some cases the data needed for a Particular study already exist. Companies maintain a variety of Data bases about their Employees, customers and business operations. Data on Employee Salaries, Qualification and Experience, age etc., can be obtained from the company records. Other internal records contain data on Sales advertising Expenditure, inventory levels etc. Today the internet has become a very important Source of data and statistical information. Almost all the Companies today maintain websites.

Statistical Studies :-

Sometimes the data needed for a particular study may not be available through existing sources. In such cases, the data can be obtained by conducting statistical studies which may be "Experimental or Observational". In Experimental study, a variable of interest is first identified and other variables are identified that influence the variable of interest.

Ex:- A pharmaceutical company might be interested in conducting an experiment to learn how a new vaccine will perform on any deadly disease. Similarly a fertilizer manufacturing company also may be interested to know the improvement in the crop yield.

Some times observational (or) non-experimental study also will be conducted by the companies to collect the data.

Ex:- Observing the consumer buying behaviour in the Marts.

Measure of central Tendency :-

A useful way to describe a set of observations is to find out a single value, that represents the average characteristics of its frequency distribution such a value is called measure of central Tendency. Since it is generally located towards the centre of the Distribution it is called Measure of central Tendency. The Three most commonly used measures of central Tendency are

1. Mean.
2. Median.
3. Mode.

Mean :-

The mean is also called Arithmetic Mean / Average / Simply Mean. The Arithmetic mean is obtained by adding all the

observations and dividing the sum with the number of observations consider the following examples.

Ex:- 10, 15, 30, 7, 42, 49, 83 the mean is calculated as follows.

$$\overline{X} = \frac{\sum x}{n}$$

Mean $\rightarrow$ Sum of observations
 $n \rightarrow$ No. of observations

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7}{7}$$

$$= \frac{10 + 15 + 30 + 7 + 42 + 49 + 83}{7}$$

$$= \frac{206}{7}$$

$$\overline{x} = 38$$

NOTE:- In case of ungrouped (Raw) data, weights are involved, the approach for calculating the arithmetic mean is as follows.

$\rightarrow$ for Examples the following are the marks obtained by a student in 3-tests, mid term test - 30

Lab - 25

Final - 20

$$\text{Simple Arithmetic Mean} = \frac{30 + 25 + 20}{3} = \frac{75}{3} = 25$$

Suppose, if weights are assigned on the basis of their weights.

Mid term - 2 points

Lab - 3 points

Final - 5 points.

Calculation of Weight Mean

Test	Relative Weight (W)	Marks (X)	WX
Mid Term	2	30	60
Laboratory	3	25	50
-final	5	20	100
	$\Sigma W = 10$		$\Sigma WX = 235$

$$\text{Weighted Mean} = \frac{\Sigma WX}{\Sigma W}$$

$$= \frac{235}{10} \Rightarrow 23.5$$

- ② An investor is fond on investing in share market in stock Exchanges. The stock is sold at Rs. 120/- per share on 1st day, Rs. 105/- on the 2nd day and Rs. 90/- on the 3rd day. The investor has purchased 50 shares on the 1st day, 80 shares on the 2nd day and 100 shares on the 3rd day. What avg price per share did the investor pay.

Soln

Day	Price Per Share (X)	No. of shares per day (W)	WX
1	120	50	6000
2	105	80	8400
3	90	100	9000
		$\Sigma W = 230$	$\Sigma WX = 23400$

$$\text{Weighted Mean} = \frac{\Sigma WX}{\Sigma W}$$

$$= \frac{23400}{230}$$

$$= 101.7$$

Grouped Data:-

For grouped data, the Arithmetic Mean can be calculated by a number of Methods. one of the popular Methods is direct Method. In Direct Method the mean is calculated using the following formula

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

Where

m = mid point of various classes

f = frequency of each class

n = Total no. of frequencies.

Ex:- The following are the marks obtained by 58 MBA students in statistics. calculate the Average marks of the students.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	4	8	11	15	12	6	2

Soln

Marks	midpoint (m)	No. of students (f)	fm
0-10	5	4	20
10-20	15	8	120
20-30	25	11	275
30-40	35	15	525
40-50	45	12	540
50-60	55	6	330
60-70	65	2	130
			1940

$$\begin{aligned}\bar{x} &= \frac{\sum fm}{n} = \frac{1940}{58} \\ &= 33.45 \\ &\approx 33.\end{aligned}$$

3) The mean of the following frequency distribution was found to be 1.46. calculate the missing frequency.

No. of Accidents	0	1	2	3	4	5	
No. of Days	46	-	-	25	10	5	200

Mean = 1.46

Soln

$$\text{Mean} = \frac{(0 \times 46) + x + 2y + (3 \times 25) + (4 \times 10) + (5 \times 5)}{200}$$

$$1.46 = \frac{x + 2y + 75 + 40 + 25}{200}$$

$$(200 \times 1.46) = x + 2y + 140$$

$$x + 2y = 292 - 140$$

$$\boxed{x + 2y = 152} \rightarrow \textcircled{1}$$

$$46 + x + y + 25 + 10 + 5 = 200$$

$$86 + x + y = 200$$

$$\boxed{x + y = 114} \rightarrow \textcircled{2}$$

Solving $\textcircled{1}$ & $\textcircled{2}$

$$\begin{array}{r} x + 2y = 152 \\ x + y = 114 \\ \hline (-) (-) \quad (-) \end{array}$$

$$y = 38$$

$$x + 2(38) = 152$$

$$x + 76 = 152$$

$$x = 76, y = 38$$

Combined Mean :-

When two related groups are given, the combined mean taking both the groups into consideration is calculated using the following formula.

$$\bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

Where $\Rightarrow \bar{x}_1$ & $\bar{x}_2$ are the Means of group 1 and group 2 respectively

$\Rightarrow n_1$ & n_2 are the no. of observations of group 1 and group 2 respectively.

$\Rightarrow \bar{x}_{12}$ is the combined mean of two groups.

① The mean height of 25 male workers in a factory is 61" inches and the mean height of 35 female workers in a factory is 58" inches. Find out the combined mean height of workers in the factory.

$$\text{Combined Mean } \bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

Given Data

$$n_1 = 25 \quad \bar{x}_1 = 61''$$

$$n_2 = 35 \quad \bar{x}_2 = 58$$

$$\bar{x}_{12} = \frac{(25 \times 61) + (35 \times 58)}{25 + 35}$$

$$= \frac{1525 + 2030}{60} \Rightarrow \frac{3555}{60} = 59.25 \approx 59''.$$

2) The mean annual salary of employees of a company is Rs. 30,000/- The mean annual salaries of male & female employees are Rs. 35,000/- and Rs. 23,000/- respectively. Find out the Percentage of male & female employee working in the company.

Given Data $\bar{x}_{12} = 30,000$, $\bar{x}_1 = 35,000$ - Males
 $\bar{x}_2 = 23,000$ - Females.

$$30,000 = \frac{35,000 n_1 + 23,000 n_2}{n_1 + n_2}$$

$$30,000 n_1 + 30,000 n_2 = 35,000 n_1 + 23,000 n_2$$

$$7,000 n_2 = 5,000 n_1$$

$$\frac{n_2}{n_1} = \frac{5}{7} \Rightarrow \boxed{\frac{n_1}{n_2} = \frac{7}{5}}$$

where n_1 = number of males.

n_2 = number of Females.

$$\therefore \text{males} = \frac{n_1}{n_1 + n_2} \times 100 = \frac{7}{7+5} \times 100 \Rightarrow \frac{7}{12} \times 100 = 58.33\%$$

$$\therefore \text{females} = \frac{n_2}{n_1 + n_2} \times 100 = \frac{5}{12} \times 100 = 41.67\%$$

14/10/17

Median:-

Ungroup:- Median is defined as the value of the middle item or the mean of the values of two middle items, when the data is arranged in ascending or descending order. Thus, in an ungrouped (raw) frequency distribution if n values are arranged in ascending or descending of their values, the median is the middle value if n is odd when n is even, the median is the mean of the two middle items.

Ex:- Consider the following Series of numbers 15, 19, 21, 7, 10, 33, 25, 18, 5
Re-Arranging the above observations in ascending order

5, 7, 10, 15, 18, 19, 21, 25, 33

Median is 18.

Suppose, in the above Series there exist one more item 23. Now re-arranging the Series once again in the ascending order.

$$5, 7, 10, 15, \boxed{18, 19}, 21, 23, 25, 33 \Rightarrow \frac{18+19}{2} = \frac{37}{2} = 18.5$$

Median is 18.5

Grouped Data:-

In case of the grouped data the median is calculated using the formula

$$M = L_1 + \left(\frac{L_2 - L_1}{f} \right) (m - c) \quad m = \frac{n+1}{2}$$

Where

M = Median

L_1 = lower limit of the class in which median lies

L_2 = upper limit of the class in which median lies

f = frequency of the class in which median lies

m = Middle item where n stands for Median.

c = Cumulative frequency of the class Preceding the Median class.

① for the following frequency distribution calculate the Median

Monthly wages (Rs)	800 - 1000	1000 - 1200	1200 - 1400	1400 - 1600	1600 - 1800	1800 - 2000	
No. of Workers	18	25	30	34	26	10	143 Total

Soln

Monthly wages	No. of Workers frequency	cumulative frequency
800 - 1000	18	18
1000 - 1200	25	43
L_1 1200 - 1400 L_2	30 f	73
1400 - 1600	34	107
1600 - 1800	26	133
1800 - 2000	10	143
	143	

$$M = \frac{n+1}{2} = \frac{143+1}{2} = \frac{144}{2} = 72$$

$$M = 1200 + \left(\frac{1400 - 1200}{30} \right) \times (72 - 43)$$

$$= 1200 + \frac{200 \times 29}{30}$$

$$= 1200 + \frac{5800}{30} \Rightarrow 1200 + 193.33$$

$$= 1393.33$$

2) for the following frequency distribution find Median

class Interval	15-25	25-35	35-45	45-55	55-65	65-75
frequency	4	11	19	14	0	2

Soln

class Intervals	frequency	cumulative frequency
15-25	4	4
25-35	11	15
35-45	19	34
45-55	14	48
55-65	0	48
65-75	2	50
	50	

$$\left(\frac{n}{2}\right)^{th} = \frac{50}{2} = 25$$

$$m = \frac{50+1}{2} = \frac{51}{2} = 25.5$$

$$M = 35 + \left(\frac{45-35}{19}\right) (25.5-15)$$

$$= 35 + \frac{10 \times 10.5}{19} = 35 + \left(\frac{105}{19}\right) = 35 + 5.526$$

$$= 40.526$$

3) calculate the Median from the following data

Value	frequency	value	frequency
less than 10	4	less than 50	96
less than 20	16	less than 60	112
less than 30	40	less than 70	120
less than 40	76	less than 80	125

Soln

class Interval	frequency	cumulative frequency
0 - 10	4	4
10 - 20	12 (16-4)	16
20 - 30	24 (40-16)	40 C
l_1 30 - 40 l_2	36 f (76-40)	76
40 - 50	20	96
50 - 60	16	112
60 - 70	8	120
70 - 80	5	125

Median class $\left(\frac{n}{2}\right)^{th}$

$$= \frac{125}{2} = 62.5$$

$$m = \frac{125+1}{2} = \frac{126}{2} = 63$$

Median $M = l_1 + \frac{(l_2 - l_1)}{f} (m - c)$

$$= 30 + \frac{(40 - 30)}{36} (63 - 40)$$

$$= 30 + \frac{10 \times 23}{36} = 30 + \frac{230}{36}$$

$$= 30 + 6.38 = \underline{36.38}$$

4) find out the Missing frequencies if $n = 100$ and Median = Rs. 50/-

Expenditure	0-20	20-40	40-60	60-80	80-100
No. of frequency	14	?	26	?	16

Soln

Assuming missing frequencies for class interval 20-40 as x and for the class interval 60-80 as y .

$$14 + x + 26 + y + 16 = 100$$

$$56 + x + y = 100 \Rightarrow \boxed{x + y = 44} \rightarrow (1)$$

Expenditure	frequency	cumulative frequency
0-20	14	14
20-40	x	14+x
40-60	26	40+x
60-80	y	40+x+y
80-100	16	56+x+y

$$\text{Median} = R_f \cdot 50/-$$

$$m = \frac{n+1}{2} = \frac{100+1}{2}$$

$$= 101/2$$

$$m = 50.5$$

$$\text{Median } M = 40 + \left(\frac{60-40}{26} \right) (50.5 - 14 - x)$$

$$50 = 40 + \left(\frac{20}{26} \right) (36.5 - x)$$

$$50 = 40 + 0.7692 (36.5 - x)$$

$$50 = 40 + 28.07 - 0.76x$$

$$50 = 68.07 - 0.76x$$

$$0.76x = 18.07$$

$$x = \frac{18.07}{0.76} = 23.78$$

$$\boxed{x = 23.78}$$

Substitute in Eq(1)

$$x + y = 44$$

$$y = 44 - 23.78$$

$$\boxed{y = 20.23}$$

MODE :-

Raw Data :-

The mode is another important measure of central tendency it is the value at the point around which the observations are most heavily concentrated as an example

Ex:- Consider the following Series 8, 9, 11, 15, 16, 12, 15, 3, 7, 15
 There are 10 observations in the Series and the observation "15" appeared for maximum no. of times that is 3 times.
 Hence, Mode = 15.

Group Data:-

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times i/c \quad i \text{ (or) } c$$

Where, l_1 = The lower value of the class in which mode lies

f_1 = frequency of the class in which mode lies.

f_0 = The frequency of the class preceding the Modal class

f_2 = The frequency of the class succeeding the modal class

c/i = The class interval of Model class.

16/10/17

class Interval	30-40	40-50	50-60	l_1 60-70	70-80	80-90	90-100
frequency	4	6	8 f_0	12 f_1	9 f_2	7	4

Soln

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times c$$

$$= 60 + \frac{12 - 8}{(12 - 8) + (12 - 9)} \times 10$$

$$= 60 + \frac{4}{4 + 3} \times 10$$

$$= 60 + \frac{4}{7} \times 10$$

$$= 60 + \frac{40}{7}$$

$$= 60 + 5.7$$

$$= 65.7$$

- ① The following is the frequency distribution of weights of 60 students find the Modal value

Weighting	No. of students
30 - 35	3
35 - 40	8
40 - 45	12 $- f_0$
45 - 50	20 $- f_1$
50 - 55	15 $- f_2$
55 - 60	2

Soln

$$M = 45 + \frac{20 - 12}{(20 - 12) + (20 - 15)} \times 5$$

$$= 45 + \frac{8}{8 + 5} \times 5 \Rightarrow 45 + \frac{8 \times 5}{13}$$

$$= 45 + \frac{40}{13} \Rightarrow 45 + 3.0769$$

$$= 48.0769$$

- 2) The Expenditure details of 1000 families are given below

Expenditure in Rs.	40 - 59	60 - 79	80 - 99	100 - 119	120 - 139
No. of families	50	?	500	?	50

The median for the distribution is Rs. 87 calculate the missing frequencies and also calculate the Mode for the total distribution

Soln

Let the Missing frequencies be x & y .

Expenditure (Rs)	No. of families	Cumulative frequency
40 - 59	50	50
60 - 79	x	50 + x c
80 - 99	500	550 + x M
100 - 119	y	550 + x + y
120 - 139	50	600 + x + y
	Total = 1000	

$$50 + x + 500 + y + 50 = 1000$$

$$600 + x + y = 1000 \Rightarrow x + y = 400 \rightarrow \textcircled{1}$$

$$m = \frac{n+1}{2} = \frac{1001}{2} = 500.5$$

$$\text{Median (M)} = l_1 + \frac{l_2 - l_1}{f} + (m - c)$$

$$87 = 80 + \frac{99 - 80}{500} \times (500.5 - (50 + x))$$

$$87 = 80 + \frac{19}{500} \times (500.5 - 50 - x)$$

$$87 = 80 + \underbrace{(0.038)}_{\approx 0.04} \times (450.5 - x)$$

$$87 = 80 + 0.04 \times 450.5 - 0.04x$$

$$0.04x = -7 + 18.02 \Rightarrow 0.04x = 11.02$$

$$x = \frac{11.02}{0.04} = 275.5 \approx 276$$

$$x = 276$$

Eq $\rightarrow \textcircled{1}$

$$x + y = 400 \Rightarrow 276 + y = 400$$

$$y = 124$$

ii)

Expenditure	No. of families
40 - 59	50
60 - 79	276 f_0
80 - 99	500 f_1
100 - 119	124 f_2
120 - 139	50
	1000

$$\text{Mode (M)} = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times c$$

$$= 80 + \frac{500 - 276}{(500 - 276) + (500 - 124)} \times 19$$

$$= 80 + \frac{224}{224 + 376} \times 19$$

$$= 80 + \frac{224}{600} \times 19$$

$$= 80 + 7.03$$

$$= 87.03$$

NOTE:- The formula which is relating the mean, median & mode

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

MEASURES OF DISPERSION

The Measure of central Tendency will indicate the central value of the distribution. However, the central value alone is not sufficient to fully describe the distribution. In addition to the measures of the spread of the actual data. The extent of such spread may vary from one distribution to another distribution.

Example:-

Consider two Income distributions, they may have identical mean but they may present different patterns of distribution i.e., one distribution may have less variability than other distribution. Such variability is measured by the "Measures of Dispersion".

→ A study of dispersion or Variability enables us to get additional information about the composition of data.

It is important to note that a high degree of uniformity means low degree of dispersion.

There are four Measures of Dispersion

1. RANGE
2. QUARTILE DEVIATION.
3. MEAN DEVIATION.
4. STANDARD DEVIATION.

1. The RANGE:-

Ungrouped Data:- The Simplest measure of dispersion is "The Range" which is the difference between the maximum value and the minimum value".

Ex:- Consider the marks obtained by 10 students in an examination
as 65, 68, 76, 92, 72, 57, 51, 79, 81.

$$\text{Range} = \text{Max} - \text{Min}$$

$$= 94 - 51$$

$$= 43$$

Grouped Data:-

In a frequency distribution, the range is calculated taking the difference between the upper limit of the highest class and lower limit of the lowest class.

Size of item	20-40	40-60	60-80	80-100	100-120	
frequency	7	11	30	17	5	70

Soln

$$\text{Highest class} = 100 - 120$$

$$\text{Lowest class} = 20 - 40$$

$$\text{Range} = 120 - 20 = 100$$

co-efficient of Range:-

The co-efficient of Range is calculated using the formula

$$\frac{L - S}{L + S}$$

Ex:- for the data given below 8, 10, 20, 9, 15, 10, 13, 28

$$\text{co-efficient of Range} = \frac{L - S}{L + S}$$

$$= \frac{28 - 8}{28 + 8}$$

$$= \frac{20}{36} = 0.55$$

2) The QUARTILE DEVIATION:-

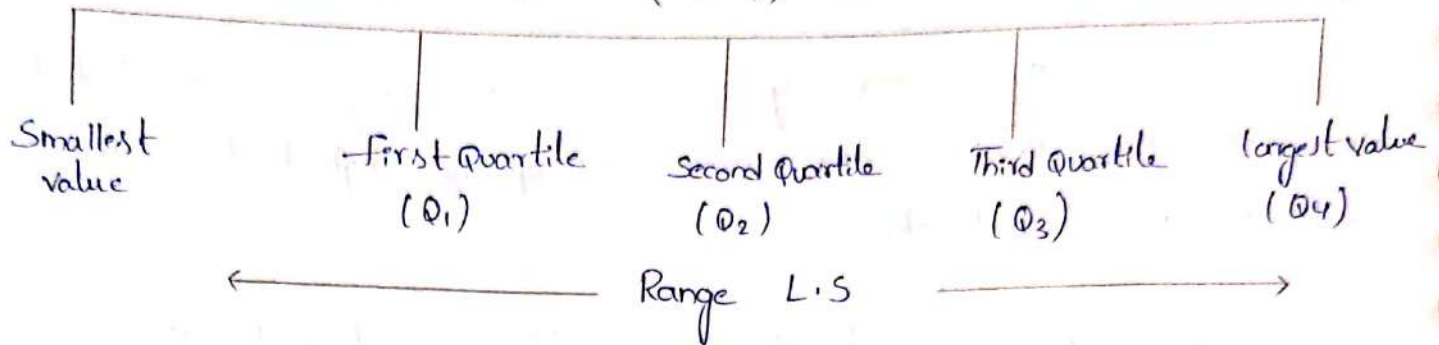
The Quartile Deviation is better measure of variation in a distribution than range. Here the middle 50% of the distribution is used by avoiding the 25% of the distribution at both the ends. In other words, the inter Quartile range or Quartile deviation denotes the difference between the 3rd Quartile and the first Quartile.

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

Where

Q_1 = Size of $(n/4)^{th}$ item.

Q_3 = Size of $(3(n/4))^{th}$ item.



In other words, the inter quartile range (or) deviation. The range for the middle 50% of the data. Quartiles are the points which divide the data into four equal parts. ' Q_1 ' distribution and ' Q_3 ' the value of the item $(3/4)^{th}$ the way up in the distribution. 3/4 Q_1 & Q_3 are included half the total number of items.

1) calculate the quartile deviation from the following data

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
frequency	11	18	25	28	30	33	22	15	22

Soln

Marks	frequency	C.F
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182
80-90	22	204
	<u>204</u>	

$$Q_1 = \text{Size of } \left(\frac{N}{4}\right)^{\text{th}} \text{ item} = \frac{204}{4} = 51^{\text{st}} \text{ item}$$

The 51st item lies in the class interval 20-30

$$Q_1 = l_1 + \frac{\left(\frac{N}{4}\right) - c}{f} \times i$$

$$= 20 + \frac{51 - 29}{25} \times 10$$

$$= 20 + \frac{22}{25} \times 10$$

$$= 20 + 8.8$$

$$= 28.8.$$

$$Q_3 = \text{Size of } \left(\frac{3N}{4}\right)^{\text{th}} \text{ item} = \left(\frac{3 \times 204}{4}\right) = 153^{\text{rd}} \text{ item.}$$

Which lies in the class interval 60-70

$$Q_3 = l_1 + \frac{\left(\frac{3N}{4}\right) - c}{f} \times i$$

$$= 60 + \frac{153 - 145}{22} \times 10$$

$$= 60 + \frac{8}{22} \times 10$$

$$= 60 + 3.63$$

$$= 63.63.$$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{63.63 - 28.8}{2} = 17.42 \text{ marks.}$$

27 Consider the following example to calculate the Quartile Deviation

class interval	30-40	40-50	50-60	60-70	70-80	80-90	90-100
frequency	4	6	8	12	9	7	4

class Interval	frequency	c.f
30 - 40	4	4
40 - 50	6	10
50 - 60	8	18
60 - 70	12	30
70 - 80	9	39
80 - 90	7	46
90 - 100	4	50
	50	

$$Q_1 = \text{Size of } \left(\frac{N}{4}\right)^{\text{th}} \text{ item} = \frac{50}{4} = 12.5^{\text{th}} \text{ item}$$

the 15th item lies in the c.I 50 - 60

$$\begin{aligned}
 Q_1 &= l + \frac{\left(\frac{N}{4}\right) - c}{f} \times i \\
 &= 50 + \frac{12.5 - 10}{8} \times 10 \\
 &= 50 + \frac{2.5}{8} \times 10 \\
 &= 50 + 0.3125 \times 10 \\
 &= 53.125
 \end{aligned}$$

$$Q_3 = \text{Size of } \left(\frac{3N}{4}\right)^{\text{th}} \text{ item} = \frac{3 \times 50}{4} = \frac{150}{4} = 37.5^{\text{th}} \text{ item}$$

$$\begin{aligned}
 Q_3 &= l + \frac{\left(\frac{3N}{4}\right) - c}{f} \times i \\
 &= 70 + \frac{37.5 - 30}{9} \times 10 \\
 &= 70 + 0.833 \times 10 \Rightarrow 78.33
 \end{aligned}$$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{78.33 - 53.12}{2} = 12.6042$$

* The coefficient of Quartile Deviation is calculated by using the formula

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$\Rightarrow \frac{78.333 - 53.125}{78.333 + 53.125} = \frac{25.2080}{131.4580} = 0.1918.$$

3) The MEAN DEVIATION:-

The mean Deviation is also known as the Average deviation. As the name implies, it is the average of the absolute amounts by which the individual items deviate from the mean. Since the positive deviations from the mean are equal to the negative deviations, while computing the mean deviation we ignore the positive and Negative Signs.

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

where M.D = The mean absolute deviation.

$|x|$ = The absolute deviation of each score from the ignoring $+/-$ signs.

n = Total no. of scores.

① consider the following Series 4, 6, 8, 10, 12 calculate Mean Deviation

$$\text{Mean} = \frac{4+6+8+10+12}{5} = \frac{40}{5} = 8.$$

The deviation are calculated as follows

$$4 - 8 = -4$$

$$6 - 8 = -2$$

$$8 - 8 = 0$$

$$10 - 8 = 2$$

$$12 - 8 = 4$$

Ignoring the Positive and Negative Signs

Sum of deviations = 12

$$\therefore \text{Mean Deviation} = \frac{\sum |x|}{n} = \frac{12}{5} = 2.4$$

Q) Consider the following frequency distribution and find the mean Deviation

Size of item frequency

2-4 20

4-6 40

6-8 30

8-10 10

100

$$\text{mean } (\bar{x}) = \frac{\sum mf}{\sum f}$$

$$\bar{x} = \frac{560}{100} = 5.6$$

Size	f	mid point (m)	mf	d from $\bar{x}$	f · d
2-4	20	3	60	-2.6	52
4-6	40	5	200	-0.6	24
6-8	30	7	210	1.4	42
8-10	10	9	90	3.4	34
	<u>100</u>		<u>560</u>		<u>152</u>

$$\text{Mean deviation} = \frac{\sum f|d|}{\sum f} = \frac{152}{100} = 1.52.$$

Standard Deviation (S.D):-

The fourth measure of dispersion is the Standard deviation. It is similar to the Mean deviation and here the deviations are measured from the mean at the same time, the Standard deviation is considered as a more desirable method, compared to the Mean Deviation.

The S.D is represented by " σ ". The square of S.D (σ^2) is called Variance.

Q) calculate the S.D from the following data 20, 15, 19, 24, 16, 14

$$\text{Mean } \mu = \frac{20+15+19+24+16+14}{6}$$

$$= \frac{108}{6} = 18$$

x	$x - \mu$	$(x - \mu)^2$
20	2	4
15	-3	9
19	1	1
24	6	36
16	-2	4
14	-4	16
		<u>70</u>

$$S.D(\sigma) = \sqrt{\frac{\sum (x - \mu)^2}{n}}$$

$$= \sqrt{\frac{70}{6}} = \sqrt{11.6} = 3.41$$

Also the S.D can be calculated with the following formula

$$\sigma = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n}}$$

x	x_i^2
20	400
15	225
19	361
24	576
16	256
14	196
108	2014

$$= \sqrt{\frac{2014 - \frac{(108)^2}{6}}{6}}$$

$$= \sqrt{\frac{2014 - 1944}{6}}$$

$$\sigma = \sqrt{11.66}$$

$$\sigma = 3.41$$

Grouped Data:-

The formula for calculating the S.D for Grouped Data

$$\sigma = C \cdot \sqrt{\frac{\sum fd_i^2}{N} - \left(\frac{\sum fd_i}{N} \right)^2}$$

1) consider the following distribution related to the Marks obtained by students in an examination.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No of Students	1	3	6	10	12	11	6	3	2	1

Marks	No. of Students (f)	mid point (m)	Deviation d	$d_i = d/10$	$-f d_i$	$-f d_i^2$	d_i^2
0 - 10	1	5	-45	-4.5	-4.5	20.25	20.25
10 - 20	3	15	-35	-3.5	-10.5	36.75	12.25
20 - 30	6	25	-25	-2.5	-15.0	37.5	6.25
30 - 40	10	35	-15	-1.5	-15	22.5	2.25
40 - 50	12	45	-5	-0.5	-6	30	0.25
50 - 60	11	55	5	0.5	5.5	2.75	0.25
60 - 70	6	65	15	1.5	9	13.25	2.25
70 - 80	3	75	25	2.5	7.5	18.25	6.25
80 - 90	2	85	35	3.5	7.0	24.5	12.25
90 - 100	1	95	45	4.5	4.5	20.25	20.25
	$n = 55$	$\Sigma m = 500$			$\Sigma -f d_i = -17.5$	$\Sigma -f d_i^2 = 199.5$	

$$\text{Mean} = \frac{500}{10} = 50$$

$$\begin{aligned}
 \therefore \sigma &= C \cdot \sqrt{\frac{\Sigma f d_i^2}{N} - \left(\frac{\Sigma f d_i}{N}\right)^2} \\
 &= 10 \sqrt{\frac{199.5}{55} - \left(\frac{-17.5}{55}\right)^2} \\
 &= 10 \sqrt{3.63 - 0.1024} \\
 &= 10 \sqrt{3.53} \\
 &= 10(1.88) \\
 &= 18.8
 \end{aligned}$$

*** COMBINED STANDARD DEVIATION:-

The formula for calculating the combined S.D of two groups containing n_1 and n_2 observations with means $\bar{x}_1$ and $\bar{x}_2$ and standard deviations σ_1 and σ_2 respectively is

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Where σ_{12} = combined S.D of two groups.

$$d_1 = x_{12} - \bar{x}_1$$

$$d_2 = x_{12} - \bar{x}_2$$

$$\text{Where } x_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

↓
Combined Mean

1) The mean and S.D's of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the Mean and S.D of all the 250 items put together.

Soln

Given $n_1 = 100$, $n_2 = 150$

$$\bar{x}_1 = 50, \bar{x}_2 = 40$$

$$\sigma_1 = 5, \sigma_2 = 6$$

$$\text{Combined mean } x_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

$$= \frac{(100)(50) + (150)(40)}{100 + 150}$$

$$= \frac{5000 + 6000}{250} = \frac{11000}{250} = 44$$

$$\therefore x_{12} = 44$$

$$\Rightarrow d_1 = x_{12} - \bar{x}_1 = 44 - 50 = -6$$

$$d_2 = x_{12} - \bar{x}_2 = 44 - 40 = 4$$

(2) combined S.D (σ_{12})

$$\sigma_{12} = \sqrt{\frac{100(5^2 + (-6)^2) + 150(6^2 + (4)^2)}{100 + 150}}$$

$$\begin{aligned}
 &= \sqrt{\frac{100(25+36) + 150(36+16)}{250}} \\
 &= \sqrt{\frac{6100 + 7800}{250}} \Rightarrow \sqrt{\frac{13900}{250}} \\
 &= \sqrt{55.6} \\
 &= 7.46
 \end{aligned}$$

Co-efficient of Variation (C.V):-

The S.D is an absolute measure of dispersion as it measures the variation in the same units as the original data. As search the S.D be a suitable measure while comparing two or more distributions. In ^{Search} such cases a relative measure of dispersion is used which is called Co-efficient of Variation.

The formula for calculating the co-efficient of Variation.

$$\boxed{CV = \frac{\sigma}{\mu} \times 100} \quad \text{where } \sigma = \text{S.D} \\
 \mu = \text{Mean}$$

- 1) A Small business for these are two types A, B. Typist 'A' on an average types 30 pages with S.D of 6. Typist 'B' types on an average 45 pages for per day with S.D of 10. Which typist shows greater consistency in their outputs

Soln

$$CV \text{ for typist A} = \frac{\sigma}{\mu} \times 100$$

$$= \frac{6}{30} \times 100$$

$$= 20\%$$

$$CV \text{ for typist B} = \frac{\sigma}{\mu} \times 100$$

$$= \frac{10}{45} \times 100$$

$$= 22.22\%$$

The above calculations clearly indicate that although typist 'B' types more pages, there is a greater variation in his output compared to that of typist 'A'. Hence typist 'A' is more consistency.

2) findout the Missing information in the following Table.

	A	B	C	Combined
Number	50	60	90	200
SD	6	7	8	7.746
Mean	113	120	115	116

① Number: $50 + x + 90 = 200$

$$\Rightarrow x + 140 = 200 \Rightarrow x = 60$$

$$\text{Combined Mean} = \bar{x}_{123} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3}{n_1 + n_2 + n_3}$$

$$116 = \frac{(50 \times 113) + (60 \times y) + (90 \times 115)}{50 + 60 + 90}$$

$$116 = \frac{5650 + 60y + 10350}{200}$$

$$23200 = 16000 + 60y$$

$$23200 - 16000 = 60y$$

$$7200 = 60y \Rightarrow y = 120$$

$$\sigma_{123} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + n_3(\sigma_3^2 + d_3^2)}{n_1 + n_2 + n_3}}$$

$$d_1 = \bar{x}_{123} - \bar{x}_1$$

$$= 116 - 113$$

$$= 3$$

$$d_2 = \bar{x}_{123} - \bar{x}_2$$

$$= 116 - 120$$

$$= -4$$

$$d_3 = \bar{x}_{123} - \bar{x}_3$$

$$= 116 - 115$$

$$= 1$$

$$7.746 = \sqrt{\frac{50(6^2 + 3^2) + 60(7^2 + (-4)^2) + 90(8^2 + 1^2)}{50 + 60 + 90}}$$

$$7.746 = \sqrt{\frac{50(36+9) + 60(49+16) + 90(z^2+1)}{200}}$$

$$7.746 = \sqrt{\frac{2250 + 3900 + 90z^2 + 90}{200}}$$

$$7.746 = \sqrt{\frac{6240 + 90z^2}{200}}$$

$$(7.746)^2 = \frac{6240 + 90z^2}{200}$$

$$60 = \frac{6240 + 90z^2}{200}$$

$$12000 = 6240 + 90z^2$$

$$90z^2 = 5760$$

$$z^2 = \frac{5760}{90}$$

$$z = \sqrt{64} \Rightarrow z = 8$$

3)

Bursting Pressure	5-10	10-15	15-20	20-25	25-30
No. of Bag's A	2	9	29	54	6
No. of Bag's B	9	15	30	32	4

From these two samples determine which Manufacturer's bags have higher average bursting pressure and which of them is more uniform in the bursting pressure.

Bag A : Co-efficient of Variation (cv) = $\frac{\sigma}{\mu} \times 100$

Mean

C.I	Bag A (f)	Midpoint (m)	-fm
5-10	2	7.5	15
10-15	9	12.5	112.5
15-20	29	17.5	507.5
20-25	54	22.5	1215
25-30	6	27.5	165
	100		2015

$$\begin{aligned} \text{Mean} &= \frac{\sum fm}{N} \\ (\bar{x}) &= \frac{2015}{100} \\ &= 20.15 \end{aligned}$$

S.D

C.I	f	Midpoint (m)	d	d ²	d ²	fd ²	fd ²
5-10	2	7.5	-10	-2	4	-4	8
10-15	9	12.5	-5	-1	1	-9	9
15-20	29	17.5	0	0	0	0	0
20-25	54	22.5	5	1	1	54	54
25-30	6	27.5	10	2	4	12	24
	100					53	95

$$\begin{aligned} d &= m - \frac{\sum m}{N} \\ &= \frac{100}{100} \end{aligned}$$

$$\sigma = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= \sqrt{\frac{95}{100} - \left(\frac{53}{100}\right)^2}$$

$$= \sqrt{0.95 - 0.28}$$

$$= \sqrt{0.67}$$

$$= 5(0.82) = 4.1$$

$$d' = \frac{d}{\text{difference b/w class interval}}$$

$$C.V = \frac{\sigma}{\mu} \times 100$$

$$= \frac{4.1}{20.15} \times 100 = 20.34$$

Bag B:

Co-efficient of variation $CV = \frac{\sigma}{\mu} \times 100$

Mean

C.I	frequency (f)	Midpoint (m)	-fm
5-10	9	7.5	67.5
10-15	15	12.5	187.5
15-20	30	17.5	525
20-25	32	22.5	720
25-30	4	27.5	110
	90		1610

$$\text{Mean} = \frac{\sum fm}{N}$$

$$\mu = \frac{1610}{90}$$

$$\mu = 17.88$$

S.D

C.I	f	Midpoint (m)	d	d'	d' ²	fd'	fd' ²
5-10	9	7.5	-10	-2	4	-18	36
10-15	15	12.5	-5	-1	1	-15	15
15-20	30	17.5	0	0	0	0	0
20-25	32	22.5	5	1	1	32	32
25-30	4	27.5	10	2	4	8	16
	90					7	99

$$\sigma = \sqrt{\left(\frac{\sum fd'^2}{N} \right) - \left(\frac{\sum fd'}{N} \right)^2}$$

$$= 5 \sqrt{\frac{99}{90} - \left(\frac{7}{90} \right)^2} \Rightarrow 5 \sqrt{1.1 - 0.0049}$$

$$\Rightarrow 5 \sqrt{1.095} \Rightarrow 5(1.046)$$

$$= 5.23$$

$$C.V = \frac{\sigma}{\mu} \times 100 \Rightarrow \frac{5.23}{17.88} \times 100 \Rightarrow 29.25$$

We observe that the average bursting pressure of bag A is more than that of bag B.

$\therefore$ The co-efficient of variation of bag A is less than that of bag B, the uniformly bursting pressure is more for bag A.

4) for the data given below, find out which of the two Series is more consistent

Variable	10-20	20-30	30-40	40-50	50-60	60-70
Series A	10	16	34	38	24	18
Series B	18	22	38	34	20	8

Soln

Series A : Mean

C.I	frequency (f)	Midpoint (m)	f.m
10-20	10	15	150
20-30	16	25	400
30-40	34	35	1190
40-50	38	45	1710
50-60	24	55	1320
60-70	18	65	1170
	140		5940

$$\begin{aligned}\text{Mean } \mu &= \frac{\sum fm}{N} \\ &= \frac{5940}{140} \\ &= 42.42\end{aligned}$$

S.D:-

C.I	f	Midpoint (M)	d	d'	d' ²	fd'	fd' ²
10-20	10	15	9	-0.9	0.81	9	8.1
20-30	16	25	-1	-0.1	0.01	1.6	0.16
30-40	34	35	-11	-1.1	1.21	-37.4	41.14
40-50	38	45	-2	-2.1	4.41	-79.8	167.58
50-60	24	55	-3	-3.1	9.61	-77.5	230.64
60-70	18	65	-41	-4.1	16.81	-73.8	302.58
	140					-257.9	750.2

$$\sigma = C \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd'}{N}\right)^2}$$

$$= 10 \sqrt{\frac{750.2}{140} - \left(\frac{-257.9}{140}\right)^2}$$

$$= 10 \sqrt{5.35 - 3.38} \Rightarrow 10 \sqrt{1.97} \Rightarrow 10(1.40) = 14.$$

$$C.V = \frac{\sigma}{\mu} \times 100$$

$$= \frac{14}{42.42} \times 100 \Rightarrow 0.33 \times 100$$

$$= 33\%$$

Series B :- Mean

C.I	frequency	Mid point	f.m
10-20	18	15	270
20-30	22	25	330
30-40	38	35	1330
40-50	34	45	1530
50-60	20	55	1100
60-70	28	65	1820
	160		6380

$$\begin{aligned} \text{Mean } \mu &= \frac{\sum fm}{N} \\ &= \frac{6380}{160} \\ &= 39.87 \end{aligned}$$

S.D:-

$$d = \frac{240}{10} = 24$$

C.I	f	Midpoint	d	d'	d ²	fd'	fd ²
10-20	18	15	9	0.9	0.81	16.2	14.58
20-30	22	25	-1	-0.1	0.01	-2.2	0.22
30-40	38	35	-11	-1.1	1.21	-41.8	45.98
40-50	34	45	-21	-2.1	4.41	-71.4	149.94
50-60	20	55	-31	-3.1	9.61	-62	192.2
60-70	28	65	-41	-4.1	16.81	-114.8	470.68
	160					-276	873.6

$$S.D \quad \sigma = C \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd'}{N} \right)^2}$$

$$= 10 \sqrt{\frac{873.6}{160} - \left(\frac{-276}{160} \right)^2}$$

$$= 10 \sqrt{5.46 - 2.975}$$

$$= 10 \sqrt{2.49}$$

$$\sigma = 10 (1.577) = 15.77$$

$$\begin{aligned} C.V &= \frac{\sigma}{\mu} \times 100 = \frac{15.77}{39.87} \times 100 \Rightarrow 0.395 \times 100 \\ &= 39.55\% \end{aligned}$$

5) The arithmetic Mean and S.D of a Series of 20 items were computed as 20cm, 5cm respectively while calculating this an item 13 was mislead as 30. find the correct mean and Standard deviation.

Soln

$$\bar{x} = 20 \quad \text{we have}$$

$$n = 20 \quad \text{values} \quad \bar{x} = \frac{\sum x}{n}$$

$$20 = \frac{\sum x}{20}$$

$$\sum x = 400$$

$$\begin{aligned} \text{Corrected } \sum x &= 400 - 30 + 13 \\ &= 383. \end{aligned}$$

$$\begin{aligned} \text{Corrected Mean} &= \frac{\text{corrected } \sum x}{n} \\ &= \frac{383}{20} = 19.15 \end{aligned}$$

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

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

$$25 = \frac{\sum x^2}{n} - \left(\frac{400}{20}\right)^2$$

$$25 = \frac{\sum x^2}{n} - \left(\frac{400}{20}\right)^2$$

$$25 = \frac{\sum x^2}{20} - 400$$

$$25 = \frac{\sum x^2 - 8000}{20}$$

$$500 = \sum x^2 - 8000$$

$$\sum x^2 = 8500$$

$$\text{Corrected } \sum x^2 = 8500 - 30^2 + 13^2$$

$$= 8500 - 900 + 169$$

$$\text{Corrected } \sum x^2 = 7769.$$

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

= $\sqrt{\frac{7769}{20} - \left(\frac{383}{20}\right)^2}$

= $\sqrt{388.45 - 366.72}$

= $\sqrt{21.72}$

= 4.660

Friday
20/10/17

UNIT - II

CORRELATION AND REGRESSION

Correlation:-

Correlation is a statistical tool which studies the relationship between two variables and correlation analysis involves various methods and techniques used for studying and measuring the extent of relationship between the two variables.

Definition:-

1. When the relationship is of Quantitative Nature, the appropriate statistical tool for measuring the relationship and expressing it in a brief formula is called CORRELATION.
2. Correlation Analysis contributes to the understanding of economic behaviour, helps in locating the critically important variables which are either dependent or independent.

NOTE:-

Two variables are said to be correlated if the change in one variable results in a corresponding change in the other variable.

* Correlation is of the following types

a) POSITIVE (OR) NEGATIVE.

b) LINEAR (OR) NON-LINEAR.

POSITIVE CORRELATION:-

If the values of two variables deviate in the same direction i.e., if there is an increase in the value of one variable which results in the increase in the value of another variable or if a decrease in the value of one variable results in the decrease of the value of another variable, the correlation is said to be POSITIVE OR DIRECT CORRELATION.

Examples:-

1. The family income and expenditure on luxury items.
2. Amount of rain-fall and crop yield.

NEGATIVE CORRELATION:-

Correlation is said to be negative or inverse if the variables deviate in the opposite direction i.e., if the decrease (increase) in the value of one variable results in a corresponding increase (decrease) in the value of another variable.

Exi- 1. Price and Demand of a commodity.

2. The sale of woollen garments and Temperature.

LINEAR CORRELATION:-

If the extent of the change in one variable tends to have a constant ratio in the extent of the change in another variable, the correlation is said to be LINEAR.

Exi-

x	5	10	15	20	25
y	30	60	90	120	150

In the above example we observe that the extent of the increase in one variable 'x' is equal to the extent of increase in the other variable 'y'.

NON-LINEAR CORRELATION:-

In Non-Linear correlation, the consistency of change does not exist.

MULTIPLE CORRELATION:-

If only two variables are involved in a study, the correlation is said to be Simple correlation. When three or more variables are involved in a study, then it is called Multiple correlation (OR) PARTIAL CORRELATION.

Karl Pearson's Method :-

According to Karl Pearson's method, The coefficient of correlation can be calculated using direct and shortcut Method.

DIRECT METHOD:- In Direct Method, The co-efficient of correlation is denoted by the symbol ' r ' which is calculated as the ratio of the co-variances of the two variables to the product of their Standard Deviations

$$r = \frac{\sum xy}{N \sigma_x \sigma_y}$$

$$\text{where } x = x - \bar{x}$$

$$y = y - \bar{y}$$

$\sum xy$ = Sum of the Product of deviations in x & y in a Series Sources from their Arithmetic Means

σ_x = Standard Deviation of Series x .

σ_y = Standard Deviation of Series y .

N = Total NO. of pairs of observations.

$$r = \frac{\sum xy}{N \sqrt{\frac{\sum x^2}{N}} \sqrt{\frac{\sum y^2}{N}}}$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}}$$

DIRECT METHOD:-

Step 1 :- Calculate the Means of the two Series $\bar{x}$ & $\bar{y}$

Step 2 :- Take the Deviation in the two Series from their respective means, indicated as x and y .

Step 3 :- Square the Deviation in both the Series and obtain the Sum of the Deviation Squared columns which gives us $\sum x^2$ and $\sum y^2$.

Step-4:- Take the product of the Deviations and add Σxy

Step-5:- The above values are used in the formula for calculating Co-efficient of correlation.

Illustration:-

X : 2 3 4 5 6

Y : 7 9 10 14 15

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
2	7	-2	-4	4	16	8
3	9	-1	-2	1	4	2
4	10	0	-1	0	1	0
5	14	1	3	1	9	3
6	15	2	4	4	16	8
				10	46	21

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

$$= \frac{2+3+4+5+6}{5} \Rightarrow \frac{20}{5} = 4$$

$$\bar{Y} = \frac{\Sigma Y}{n}$$

$$= \frac{7+9+10+14+15}{5} \Rightarrow \frac{55}{5} = 11$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \sqrt{\Sigma y^2}}$$

$$= \frac{21}{\sqrt{10} \cdot \sqrt{46}}$$

$$= \frac{21}{(3.16)(6.78)}$$

$$= \frac{21}{21.42}$$

$$= 0.98$$

NOTE:-

* The value of the correlation co-efficient obtained above i.e.,

$r = 0.98$ indicates that the two series x and y have almost perfect positive correlation as the correlation co-efficient lies between " -1 and $+1$ ".

* If the correlation co-efficient is zero, it means that the two series are not associated with each other.

* If the correlation co-efficient is " -1 ", it is Perfect Negative correlation.

* If the correlation co-efficient is " $+1$ ", it is Perfect Positive correlation.

SHORTCUT METHOD:-

Step 1:- Choose the convenient values as assumed Means of the two series x and y .

Step 2 - The Deviations i.e., dx and dy are obtained from the assumed means.

Step 3 - Obtain the sum of the deviations $\sum dx$ and $\sum dy$.

Step 4 - Findout $\sum dx dy$ and substitute in the following formula

$$r = \frac{\sum dx dy - \frac{\sum dx \sum dy}{N}}{\sqrt{\left(\sum dx^2 - \frac{(\sum dx)^2}{N}\right) \left(\sum dy^2 - \frac{(\sum dy)^2}{N}\right)}}$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{N}\right) \left(\sum y^2 - \frac{(\sum y)^2}{N}\right)}}$$

① A company manufactures different types of Electrical Appliances which are used for houses. The company uses radio for Advertising. The following Table shows the amount of time (x) in minutes and the number of Electrical Appliances sold (y) over the last Six (6) weeks

x : 25 18 32 21 35 29

y : 16 11 20 15 26 28

X	Y	dx	dy	dx ²	dy ²	dx dy
25	16	0	-4	0	16	0
18	11	-7	-9	49	81	63
32	20	7	0	49	0	0
21	15	-4	-5	16	25	20
35	26	10	6	100	36	60
29	28	4	8	16	64	32
		$\Sigma dx = 10$	$\Sigma dy = -4$	$\Sigma dx^2 = 230$	$\Sigma dy^2 = 222$	$\Sigma dx dy = 175$

NOTE:- Assuming '25' as the Assumed Mean for Series X.

Assuming '20' as the Assumed Mean for Series Y.

$$dx = x - \bar{x}$$

$$dy = y - \bar{y}$$

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

$$= \frac{175 - \frac{(10)(-4)}{6}}{\sqrt{\left(230 - \frac{100}{6}\right) \left(222 - \frac{16}{6}\right)}}$$

$$= \frac{175 + \frac{40}{6}}{\sqrt{(230 - 16.67)(222 - 2.67)}}$$

$$= \frac{175 + 6.67}{\sqrt{(213.33)(219.33)}} \Rightarrow \frac{181.67}{\sqrt{46789.66}}$$

$$= \frac{181.67}{216.31}$$

$$= 0.84$$

1) calculate the Karl Pearson co-efficient of correlation

X	50	50	55	60	65	65	65	60	60	50
Y	11	13	14	16	16	15	15	14	13	13

Soln

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
50	11	-8	-3	64	9	24
50	13	-8	-1	64	1	8
55	14	-3	0	9	0	0
60	16	2	2	4	4	4
65	16	7	2	49	4	14
65	15	7	1	49	1	7
65	15	7	1	49	1	7
60	14	2	0	4	0	0
60	13	2	-1	4	1	-2
50	13	-8	-1	64	1	8
580	140			360	22	70

$$\bar{X} = \frac{\sum X}{n} = \frac{580}{10} = 58$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{140}{10} = 14$$

$$\begin{aligned} \text{co-efficient of correlation } r &= \frac{\sum xy}{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}} \\ &= \frac{70}{\sqrt{360} \cdot \sqrt{22}} \\ &= \frac{70}{(18.97)(4.6)} \\ &= \frac{70}{88.97} \\ &= 0.79 \end{aligned}$$

2) calculate the Karl Pearson co-efficient of correlation for the given data

X :	21	22	23	24	25	26	27	28	29	30
Y :	40	39	39	35	34	34	34	31	28	25

-0.96

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
21	40	-4.5	6.1	20.25	37.21	-27.45
22	39	-3.5	5.1	12.25	26.01	-17.85
23	39	-2.5	5.1	6.25	26.01	-12.75
24	35	-1.5	1.1	2.25	1.21	-1.65
25	34	-0.5	0.1	0.25	0.01	-0.05
26	34	0.5	0.1	0.25	0.01	0.05
27	34	1.5	0.1	2.25	0.01	0.15
28	31	2.5	-2.9	6.25	8.41	-7.25
29	28	3.5	-5.9	12.25	34.81	-20.65
30	25	4.5	-8.9	20.25	79.21	-40.05
255	339			82.5	212.90	-127.5

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

$$= \frac{255}{10} = 25.5$$

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

$$= \frac{339}{10} = 33.9$$

$$\text{Co-efficient of correlation} = r = \frac{\sum xy}{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}}$$

$$= \frac{-127.5}{\sqrt{82.5} \cdot \sqrt{212.90}}$$

$$= \frac{-127.5}{(9.0830)(14.5911)}$$

$$= \frac{-127.5}{132.5310}$$

$$= -0.9620.$$

NOTE:-

The co-efficient of correlation is denoted by the letter 'r' and the square of the co-efficient of correlation ' r^2 ' denotes co-efficient of Determination. i.e., if $r = 0.9$, $r^2 = (0.9)^2 = 0.81$ which is called co-efficient of Determination.

RANK CORRELATION

The Rank correlation was Proposed by Spearman and is calculated in the following way. In Rank correlation, normally we have two types of Numerical Problems.

1. When actual ranks are given.
2. When ranks are not given.

When actual ranks are given, the differences of the two ranks $R_1 - R_2$ which is denoted by 'd' is calculated and $\sum d^2$ is found. The values are substituted in the following formula

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

NOTE:-

In the calculation of Spearman's Rank correlation co-efficient, when the ranks are not given i.e., when the actual data are given, the ranks are to be assigned. This can be done by taking the highest value as Rank one (and) or the lowest value as Rank one. Any of the approaches may be followed, but the same approach is to be followed for all the observations.

In case there is a tie between the Ranks for two data items, the average Rank is calculated.

1) The following is the data related to the Salesman ranks obtained in Training and the ranks based on Sales Performance. Calculate the co-efficient of Rank correlation and interpret.

Sales men	1	2	3	4	5	6	7	8	9	10
Ranks - Training	4	6	1	3	9	7	10	2	8	5
Ranks - Performance	5	8	3	1	7	6	9	2	10	4

Soln

Sales men	R_1	R_2	$d = R_1 - R_2$	d^2
1	4	5	-1	1
2	6	8	-2	4
3	1	3	-2	4
4	3	1	2	4
5	9	7	2	4
6	7	6	1	1
7	10	9	1	1
8	2	2	0	0
9	8	10	-2	4
10	5	4	1	1
				<u>24</u>

$$\begin{aligned}
 r_s &= 1 - \frac{6 \sum d^2}{N(N^2 - 1)} \\
 &= 1 - \frac{6(24)}{10(10^2 - 1)} \\
 &= 1 - \frac{6(24)}{990} \\
 &= 1 - \frac{144}{990} \\
 &= 1 - 0.145 \\
 &= 0.85
 \end{aligned}$$

Interpretation :- The correlation co-efficient 0.85 means that there is a high degree of relation between the Salesmen Training Program and the Salesmen Performance.

2) Find out the Spearman's Rank correlation co-efficient between the two kinds of Examination marks of MBA Students Performance in a college.

Students	A	B	C	D	E	F	G	H	I
Internal marks	51	63	73	46	50	60	47	36	60
External marks	49	72	74	44	58	66	50	30	35

Soln

Students	Internal marks	External marks	R_1	R_2	$d = R_1 - R_2$	d^2
A	51	49	5	6	-1	1
B	63	72	2	2	0	0
C	73	74	1	1	0	0
D	46	44	8	7	1	1
E	50	58	6	4	2	4
F	60	66	3.5	3	0.5	0.25
G	47	50	7	5	2	4
H	36	30	9	9	0	0
I	60	35	3.5	8	-4.5	20.25
						$\Sigma d^2 = 30.5$

$$R_1 = \frac{3+4}{2} = \frac{7}{2} = 3.5$$

$$\begin{aligned}
 r_s &= 1 - \frac{6 \Sigma d^2}{N(N^2-1)} \\
 &= 1 - \frac{6(30.5)}{9(9^2-1)} \\
 &= 1 - \frac{183}{720} \\
 &= 1 - 0.25 \\
 &= 0.75
 \end{aligned}$$

3) calculate the Rank correlation co-efficient between the marks in Statistics and Accountancy

Marks in Statistics	48	60	72	62	56	40	39	52	30
Marks in Accountancy	62	78	65	70	38	54	60	32	31

Marks in Stats	R ₁	Marks in Accounts	R ₂	d = R ₁ - R ₂	d ²
48	6	62	4	2	4
60	3	78	1	2	4
72	1	65	3	-2	4
62	2	70	2	0	0
56	4	38	7	-3	9
40	7	54	6	1	1
39	8	60	5	3	9
52	5	32	8	-3	9
30	9	31	9	0	0
					40

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(40)}{9(9^2-1)}$$

$$= 1 - \frac{240}{720}$$

$$= 1 - 0.33$$

$$= 0.67$$

$$r_s = 0.67$$

4) calculate the co-efficient of Rank correlation between the marks in Economics and statistics for 8 students as given below

Marks in Statistics : 15 10 20 28 12 10 16 18

Marks in Economics : 16 14 10 12 11 15 18 12

Marks in Stats	R ₁	Marks in Economics	R ₂	d = R ₁ - R ₂	d ²
15	5	16	2	3	9
10	7.5	14	4	3.5	12.25
20	2	10	8	-6	36
28	1	12	5.5	-4.5	20.25
12	6	11	7	-1	1
10	7.5	15	3	4.5	20.25
16	4	18	1	3	9
18	3	12	5.5	-2.5	6.25
					114

$$R = \frac{7+8}{2} = \frac{15}{2} = 7.5$$

$$R = \frac{5+6}{2} = \frac{11}{2} = 5.5$$

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(114)}{8(8^2-1)}$$

$$= 1 - \frac{684}{504}$$

$$= 1 - 1.3571$$

$$= -0.36$$

5) The Co-efficient of Rank correlation of the marks obtained by 10 students in Quantitative Techniques and Statistics was found to be 0.2. It was later found that the difference in the ranks in the two subjects obtained by one of the students was wrongly taken as 9 instead of 7. Find the correct value of the Rank correlation.

Soln Spearman $r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$

$$0.2 = 1 - \frac{6\sum d^2}{10(10^2-1)}$$

$$0.2 = 1 - \frac{6\sum d^2}{990}$$

$$0.2 \times 990 = 990 - 6\sum d^2$$

$$198 = 990 - 6\sum d^2$$

$$6\sum d^2 = 792$$

$$\sum d^2 = \frac{792}{6} = 132$$

$$\text{Correct } \sum d^2$$

$$= 132 - 9^2 + 7^2$$

$$= 132 - 81 + 49$$

$$= 100$$

$$\text{Corrected } r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

$$= 1 - \frac{6(100)}{990}$$

$$= 1 - 0.6061$$

$$= 0.3939$$

Monday
23/10/17

6) The following information is available. Determine the number of observations $\sum x^2 = 90$, $\sigma_y = 2.5$, $\sum xy = 60$, $r = 0.8$ where x, y are the deviations from the mean.

Sol

Given that $\sum x^2 = 90$, $\sigma_y = 2.5$, $\sum xy = 60$, $r = 0.8$

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$$

$$r = \frac{\sum xy}{n \sqrt{\frac{\sum x^2}{n}} \sqrt{\frac{\sum y^2}{n}}}$$

$$r = \frac{\sum xy}{n \cdot \sigma_x \cdot \sigma_y}$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n}}$$

$$0.8 = \frac{60}{n \sqrt{\frac{\sum x^2}{n}} \cdot 2.5}$$

Simplifying Squaring on both sides

$$(0.8)^2 = \frac{60^2}{n^2 \left(\frac{\sum x^2}{n} \right) \cdot (2.5)^2}$$

$$0.64 = \frac{3600}{n \cdot 90 \cdot (6.25)}$$

$$(0.64) = \frac{3600}{562.5n}$$

$$360n = 3600$$

$$n = 10$$

7) A person while calculating the correlation co-efficient between two variables x and y obtained the following results $n = 30$, $\sum x = 120$, $\sum x^2 = 600$, $\sum y = 90$, $\sum y^2 = 250$, $\sum xy = 356$. It was however later discovered at the time of checking that the calculations for two pairs of observations 8, 10 and 12, 7 were

wrongly entered instead of 8,12 and 10,8. Determine the correct value of correlation co-efficient

Sol

$$r = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{N} \right] \left[\sum y^2 - \frac{(\sum y)^2}{N} \right]}}$$

wrong $(\overset{x}{8}, \overset{y}{10})$ $(\overset{x}{12}, \overset{y}{7})$

correct $(\overset{x}{8}, \overset{y}{12})$ $(\overset{x}{10}, \overset{y}{8})$

corrected values

$$\sum x = 120 - 8 - 12 + 8 + 10 = 118$$

$$\sum y = 90 - 10 - 7 + 12 + 8 = 93$$

$$\sum x^2 = 600 - 8^2 - 12^2 + 8^2 + 10^2 = 556$$

$$\sum y^2 = 250 - 10^2 - 7^2 + 12^2 + 8^2 = 309$$

$$\sum xy = 356 - (8 \times 10) - (12 \times 7) + (8 \times 12) + (10 \times 8) = 368$$

$$r = \frac{368 - \frac{(118)(93)}{30}}{\sqrt{\left(556 - \frac{(118)^2}{30} \right) \left(309 - \frac{(93)^2}{30} \right)}}$$

$$r = \frac{368 - 365.8}{\sqrt{(556 - 464.13)(309 - 288.3)}}$$

$$r = \frac{2.2}{\sqrt{(91.87)(20.7)}}$$

$$r = \frac{2.2}{\sqrt{1901.7}}$$

$$r = \frac{2.2}{43.6085}$$

$$r = 0.05$$

REGRESSION ANALYSIS

Regression Analysis:-

The dictionary meaning of the word Regression is STEPPING BACK (OR) RETURNING TO THE AVERAGE VALUE. Regression Analysis in general means the estimation (OR) Prediction of the unknown value of one variable from the known value of other variable.

Regression Analysis is considered as a very important statistical tool which is extensively used in business and Economics.

Regression Analysis is a Mathematical Measure of the average relationship between two or more variables. In Regression Analysis, there exist two types of variables. The variable whose value is influenced or to be predicted is called "Dependent Variable" and the variable which influences the values (OR) is used for Prediction is called INDEPENDENT VARIABLE.

Line of Regression:-

A Line of Regression is the line which gives the best estimate of one variable for any given value of another variable. In case of two variables x and y , there will be two lines of Regression

1. y on x .

2. x on y .

NOTE:-

* The line of Regression of y on x is the line which gives the best estimate for the value of ' y ' for any specified value of ' x '. The equation of the line of Regression of y on x is

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

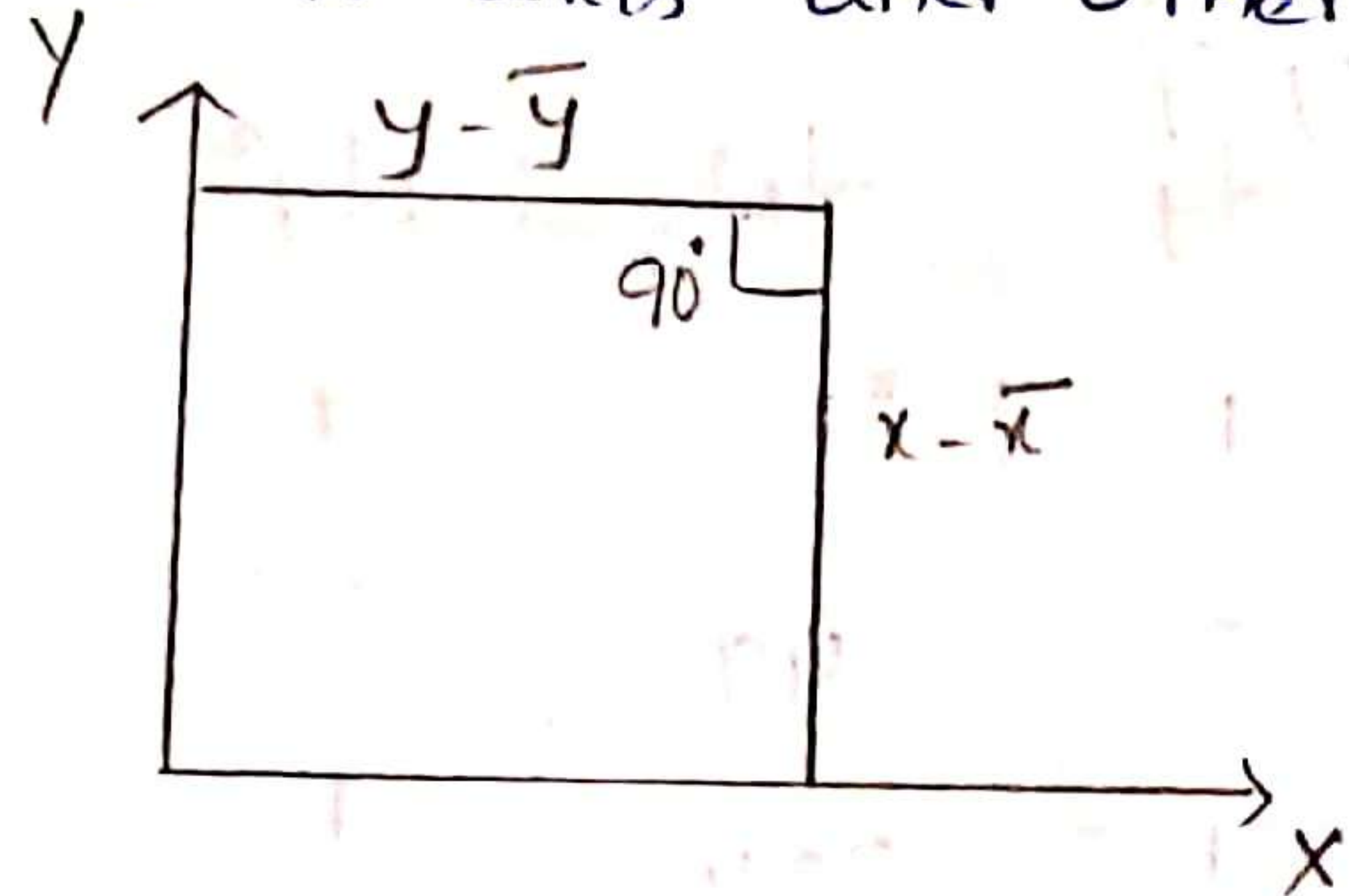
* The line of Regression of x on y is the line which gives

the best estimate for the value of x for any specified value of y . The equation of the line of Regression of x on y is

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

* The above two regression lines coincide when the correlation between the variables x and y is either perfectly positive (or) perfectly negative.

* When the correlation co-efficient is zero i.e., when the variables are uncorrelated, the two lines of Regression become Perpendicular to each other i.e., one line Parallel to x -axis and other Parallel to y -axis



* Consider the line of Regression of y on x i.e., $y = a + bx$, the co-efficient ' b ' which is the slope of the line of Regression y on x is called "co-efficient of Regression of y on x "

* For convenience Purpose for calculations, the following are taken as the co-efficients of Regression.

b_{yx} = co-efficient of Regression of y on x .

b_{xy} = co-efficient of Regression of x on y .

$\Rightarrow$

$$b_{yx} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = r \frac{\sigma_y}{\sigma_x} = \frac{\sum x dy}{\sum dx^2}$$

$$b_{xy} = \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum x dy}{\sum dy^2}$$

* The relationship between the co-efficient of correlation and the co-efficients of Regression is given by

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

the Sign to be taken before the square root is same as that of the Regression co-efficient i.e., if the Regression co-efficient is positive, consider positive sign and if the Regression co-efficient is negative consider Negative sign.

1) from the following data, obtain the two Regression Equations

Sales	91	97	108	121	67	124	51	73	111	57
Purchases	71	75	69	97	70	91	39	61	80	47

Sales x	Purchases y	$x - \bar{x}$ dx	$y - \bar{y}$ dy	dx^2	dy^2	$dx dy$
91	71	1	1	1	1	1
97	75	7	5	49	25	35
108	69	18	-1	324	1	-18
121	97	31	27	961	729	837
67	70	-23	0	529	0	0
124	91	34	21	1156	441	714
51	39	-39	-31	1521	961	1209
73	61	-17	-9	289	81	153
111	80	21	10	441	100	210
57	47	-33	-23	1089	529	759
900	700			6360	2868	3900

$$\frac{\sum x}{n} \quad \bar{x} = \frac{900}{10} = 90 \quad \bar{y} = \frac{700}{10} = 70$$

The Regression equations are as follows

$$y \text{ on } x \Rightarrow y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x}) \Rightarrow y - \bar{y} = b_{yx} (x - \bar{x})$$

$$x \text{ on } y \Rightarrow x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y}) \Rightarrow x - \bar{x} = b_{xy} (y - \bar{y})$$

$$\Rightarrow b_{yx} = \frac{\sum dx dy}{\sum dx^2} = \frac{3900}{6360} = 0.6132$$

$$\Rightarrow b_{xy} = \frac{\sum xidy}{\sum dy^2} = \frac{3900}{2868} = 1.3598$$

$$\underline{y \text{ on } x} \quad (y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$(y - 70) = 0.6132(x - 90)$$

$$\Rightarrow y - 70 = 0.6132x - 55.188$$

$$\Rightarrow \boxed{y = 0.6132x + 14.812} \text{--- (1)}$$

x on y

$$(x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$(x - 90) = 1.3598(y - 70)$$

$$x - 90 = 1.3598y - 95.1860$$

$$\Rightarrow \boxed{x = 1.3598y - 5.1860} \text{--- Regression eq. (2)}$$

The co-efficient of correlation for the above data

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

$$r^2 = b_{xy} b_{yx}$$

$$r^2 = (1.3598)(0.6132)$$

$$r^2 = 0.8338$$

$$r = \sqrt{0.8338}$$

$$\boxed{r = +0.9131}$$

2) From the data given below find out

- i) the two Regression co-efficients
- ii) The two Regression Equations
- iii) The co-efficient of correlation between the marks in Economics and Statistics
- iv) The most likely marks in Statistics when the marks in Economics are 30.

Marks in Economics : 25 28 35 32 31 36 29 38 34 32

Marks in Statistics : 43 46 49 41 36 32 31 30 33 39

Sol

∴

Economics x	Stats y	$x - \bar{x}$ dx	dy	dx^2	dy^2	$dx dy$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	380			140	398	-93

$$\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32, \quad \bar{y} = \frac{\sum y}{n} = \frac{380}{10} = 38$$

∴ The Regression Equations are as follows

$$y \text{ on } x \Rightarrow (y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$x \text{ on } y \Rightarrow (x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$\Rightarrow b_{yx} = \frac{\sum dx dy}{\sum dx^2} = \frac{-93}{140} = -0.6643$$

$$\Rightarrow b_{xy} = \frac{\sum dx dy}{\sum dy^2} = \frac{-93}{398} = -0.2337$$

y on x

$$(y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$(y - 38) = -0.6643 (x - 32)$$

$$y - 38 = -0.6643 x + 21.2576$$

$$y = -0.6643x + 59.2576$$

x on y

$$(x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$(x - 32) = (-0.2337)(y - 38)$$

$$x - 32 = -0.2337y + 8.8806$$

$$x = -0.2337y + 40.8806$$

iii) The co-efficient of correlation for the above data

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

$$r = \sqrt{(-0.6643)(-0.2337)}$$

$$r = \sqrt{0.1552}$$

$$r = -0.3940 \quad \text{Negative Regression}$$

iv) In the above calculations

$$x = 30$$

Given the marks in Economics i.e., 30, to find out the marks in Statistics, use the Regression equation on y on x

$$y - \bar{y} = b_{yx} (x - \bar{x})$$

$$(y - 38) = -0.6643(30 - 32)$$

$$y - 38 = (-0.6643)(-2)$$

$$y - 38 = 1.3286$$

$$y = 39.3286$$

3) A Departmental store use training to its salesmen which is followed by a test to terminate the salesmen whose Performance is poor. The following data gives the Test Scores and Sales made by 9 salesmen during a certain Period

Test scores	14	19	24	21	26	22	15	20	19
Sales ('00 Rs)	31	36	48	37	50	45	33	41	39

- i) Calculate the co-efficient of correlation between the Test scores and the Sales
- ii) If the firm wants a minimum Sales volume of Rs. 3000/-, what is the minimum test score needed $y =$
- iii) Also estimate the most Probable Sales volume of a salesmen making a score of 28. $x \text{ on } y$

Soln

Test Scores x	Sales y	$x - \bar{x}$ d_x	$y - \bar{y}$ d_y	d_x^2	d_y^2	$d_x d_y$
14	31	-6	-9	36	81	54
19	36	-1	-4	1	16	4
24	48	4	8	16	64	32
21	37	1	-3	1	9	-3
26	50	6	10	36	100	60
22	45	2	5	4	25	10
15	33	-5	-7	25	49	35
20	41	0	1	0	1	0
19	39	-1	-1	1	1	1
180	360			120	346	193

$$\bar{x} = \frac{\sum x}{n} = \frac{180}{9} = 20, \quad \bar{y} = \frac{\sum y}{n} = \frac{360}{9} = 40$$

The Regression equations are as follows

$$y \text{ on } x \Rightarrow (y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$x \text{ on } y \Rightarrow (x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$\Rightarrow b_{yx} = \frac{\sum dxdy}{\sum dx^2} = \frac{193}{120} = 1.6083$$

$$\Rightarrow b_{xy} = \frac{\sum dxdy}{\sum dy^2} = \frac{193}{346} = 0.5578$$

y on x

$$(y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$(y - 40) = 1.6083 (x - 20)$$

$$y - 40 = 1.6083x - 32.1660 \Rightarrow y = 1.6083x - 32.1660 + 40$$

$$\boxed{y = 1.6083x + 7.8340}$$

x on y

$$(x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$(x - 20) = 0.5578 (y - 40)$$

$$x - 20 = 0.5578y - 22.3120$$

$$\boxed{x = 0.5578y - 2.3120}$$

$$= \pm \sqrt{b_{xy} b_{yx}}$$

$$= \sqrt{(1.6083)(0.5578)}$$

$$= \sqrt{0.8971}$$

$$= 0.9472$$

ii) In the above calculations

$$y = 30$$

x on y

$$(x - 20) = b_{xy} (30 - 40)$$

$$(x - 20) = 0.5578 (-10)$$

$$x - 20 = -5.5780$$

$$\boxed{x = 14.4220}$$

iii) y on x

Given that $x = 28$

$$(y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$(y - 40) = 1.6083 (28 - 20)$$

$$y - 40 = 1.6083 (8) \Rightarrow y - 40 = 12.8664 \Rightarrow \boxed{y = 52.8664}$$

i) The co-efficient of correlation for the above data

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

$$r = \sqrt{(1.6083)(0.5578)}$$

$$r = \sqrt{0.8971}$$

$$r = +0.9472$$

Tuesday
24/10/17

4) The Data on Sales and Advertisement Expenditure of a firm is given below

	Sales	Ad. Expenditure
Mean:	40	6
Standard Deviation:	10	1.5

co-efficient of correlation $r = 0.9$

i) Estimate the likely sales for a proposed Advertisement Expenditure of Rs. 10 crores.

ii) What should be the Advertisement Expenditure if the firm Proposes a sales target of Rs. 60 crores.

Sol Here Let the variable 'x' denote the Sales in crores of Rupees and the variable 'y' denote the Advertisement Expenditure in crores of Rupees.

$$x = \text{Sales}$$

$$y = \text{Ad. Exp}$$

$$\bar{x} = 40$$

$$\bar{y} = 6$$

$$\sigma_x = 10$$

$$\sigma_y = 1.5$$

i) using the Regression Equation x on y

$$\Rightarrow x - \bar{x} = b_{xy} (y - \bar{y})$$

$$\Rightarrow x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$\Rightarrow x - 40 = 0.9 \left(\frac{10}{1.5} \right) (10 - 6)$$

$$x - 40 = 0.9 (6.67) (4)$$

$$\Rightarrow x - 40 = 24$$

$$\Rightarrow \boxed{x = 64 \text{ crores}}$$

ii) using the Regression equation y on x

$$y = 60$$

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

$$y - 6 = 0.9 \left(\frac{1.5}{10} \right) (60 - 40)$$

$$y - 6 = (0.9) (0.15) (20)$$

$$y - 6 = 2.7$$

$$\boxed{y = 8.7} \text{ crores}$$

5) When Given the following data, find out the two lines of Regression and also calculate the co-efficient of correlation

$$\Sigma x = 250$$

$$\Sigma y = 300$$

$$\Sigma xy = 7900$$

$$\Sigma x^2 = 65$$

$$\Sigma y^2 = 10000$$

$$N = 10$$

Sol

The Regression equations are

y on x

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

$$\boxed{y - \bar{y} = b_{yx} (x - \bar{x})}$$

x on y

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

$$\boxed{x - \bar{x} = b_{xy} (y - \bar{y})}$$

$$\Rightarrow b_{yx} = \frac{N \cdot \Sigma xy - \Sigma x \Sigma y}{N \Sigma x^2 - (\Sigma x)^2}$$

$$\Rightarrow b_{yx} = \frac{10(7900) - (250)(300)}{10(65) - (250)^2}$$

$$\Rightarrow b_{xy} = \frac{N \cdot \Sigma yx - \Sigma x \Sigma y}{N \cdot \Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{79000 - 75000}{650 - 62500}$$

$$= \frac{4000}{-61850} \Rightarrow r = -0.064$$

$$b_{xy} = \frac{10(7900) - (250)(300)}{10(10000) - (300)^2}$$

$$= \frac{79000 - 75000}{100000 - 90000} = \frac{4000}{10000} = 0.4$$

y on x

$$\bar{y} = \frac{\sum y}{N} = \frac{300}{10} = 30$$

$$\bar{x} = \frac{\sum x}{N} = \frac{250}{10} = 25$$

$$\hat{y} - \bar{y} = b_{yx} (x - \bar{x})$$

$$y - 30 = (-0.0647)(x - 25)$$

$$y - 30 = -0.0647x + 1.6175$$

$$y = -0.0647x + 31.6175$$

$\Rightarrow$

$$y = 31.6175 - 0.0647x$$

x on y

$$x - \bar{x} = b_{xy} (y - \bar{y})$$

$$x - 25 = 0.4 (y - 30)$$

$$x - 25 = 0.4y - 12$$

$$x = 0.4y + 13$$

$\therefore$ coefficient of correlation

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

$$r = \sqrt{(0.4)(-0.0647)}$$

$$r = \sqrt{-0.0259}$$

$$r = 0.1609$$

The square root of a negative number does not exist. for convenience purpose ignoring the negative sign

Q) By using the following data obtain the two Regression Equations

$$\bar{x} = 90 \quad \bar{y} = 70$$

$$N = 10 \quad \sum x^2 = 6360$$

$$\sum y^2 = 2860 \quad \sum xy = 3900 \quad \text{for this data}$$

Sol

Given that

$$\bar{x} = 90 \quad \bar{y} = 70$$

$$N = 10 \quad \sum x^2 = 6360$$

$$\sum y^2 = 2860 \quad \sum xy = 3900$$

The two Regression Equations are

$$x \text{ on } y \Rightarrow x - \bar{x} = b_{yx} (y - \bar{y})$$

$$y \text{ on } x \Rightarrow y - \bar{y} = b_{xy} (x - \bar{x})$$

$$b_{xy} = \frac{N \sum xy - \sum x \sum y}{N \sum y^2 - (\sum y)^2}$$

$$b_{yx} = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2}$$

$$\bar{x} = \frac{\sum x}{N} \Rightarrow 90 = \frac{\sum x}{10} \Rightarrow \sum x = 900$$

$$\bar{y} = \frac{\sum y}{N} \Rightarrow 70 = \frac{\sum y}{10} \Rightarrow \sum y = 700$$

$$b_{xy} = \frac{10(3900) - (900)(700)}{10(2860) - (700)^2}$$

$$= \frac{39000 - 630000}{28600 - 490000}$$

$$= \frac{-591000}{-461400}$$

$$b_{xy} = 1.2809$$

$$b_{yx} = \frac{10(3900) - (900)(700)}{10(6360) - (900)^2}$$

$$= \frac{39000 - 630000}{63600 - 810000} \Rightarrow \frac{-591000}{-746400} = 0.7918$$

$$b_{yx} = 0.7918$$

y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$(y - 70) = (0.7918)(x - 90)$$

$$y - 70 = 0.7918x - 71.2620$$

$$\boxed{y = 0.7918x - 1.2620}$$

x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$(x - 90) = (1.2809)(y - 70)$$

$$x - 90 = 1.2809y - 89.6630$$

$$\boxed{x = 1.2809y + 0.3370}$$

The correlation co-efficient is

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

$$r = \sqrt{(1.2809)(0.7918)}$$

$$r = \sqrt{1.0142}$$

$$r = 1.0071$$

for 100 students of a class the Regression Equation of Marks in Statistics (x) on the Marks in Commerce (y) is $3y - 5x + 180 = 0$. The mean marks in Commerce is 50 and the Variance of the Marks in Statistics is $\frac{4}{9}$ th of the Variance of the Marks in Commerce. Find the Mean Marks in Statistics and co-efficient of correlation between the Marks in the two Subjects.

Sol

$$N = 100$$

$$\text{Mean Marks in Commerce } (\bar{y}) = 50$$

Variance of Marks in Statistics is $\frac{4}{9}$ th of the Variance of Marks in Commerce

$$\boxed{\sigma_x^2 = \frac{4}{9} \sigma_y^2} \Rightarrow$$

$$\Rightarrow \frac{\sigma_x^2}{\sigma_y^2} = \frac{4}{9}$$

$$\Rightarrow \frac{\sigma_x}{\sigma_y} = \sqrt{\frac{4}{9}}$$

$$\Rightarrow \frac{\sigma_x}{\sigma_y} = \frac{2}{3}$$

Given the Line of Regression of x on y is given by

$$3y - 5x + 180 = 0$$

$$\Rightarrow -5x = -3y - 180$$

$$\Rightarrow 5x = 3y + 180$$

$$\Rightarrow 5x = 3y + 180$$

$$x = \frac{3}{5}y + \frac{180}{5}$$

$$\boxed{x = \frac{3}{5}y + 36}$$

As the Lines of Regression always pass through the point $(\bar{x}, \bar{y})$, the above Equation becomes $\bar{x} = \frac{3}{5}\bar{y} + 36$

$$= \bar{y} \quad \bar{x} = \frac{3}{5} \times 50 + 136$$

$$\bar{x} = 30 + 136$$

$$\boxed{\bar{x} = 66}$$

Correlation co-efficient

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

$$\frac{3}{5} = r \frac{2}{3}$$

$$r = \frac{3}{5} \times \frac{3}{2}$$

$$r = \frac{9}{10}$$

$$\boxed{r = 0.9}$$

8) The following is an estimated Supply Regression Equation for Sugar.

$Y = 0.025 + 1.5X$ where y is the Supply of Sugar in Kg and x is the Price of Sugar in Rupees Per Kilogram

- i) Interpret the co-efficient of the variable x
- ii) Predict the Supply when the Price is Rs. 20 per kg.
- iii) Given that $r(x, y) = 1$, in the above case, Interpret the implied relationship between Price of Sugar and Supply.

Sol

i) The Regression Equation of y (Supply in Kgs) on x (Price in Rs/kg) is given as $y = 0.025 + 1.5x$ is of the form $y = a + bx$.

We understand that the Regression co-efficient of y on x is $\boxed{b = 1.5}$ which indicates that the unit change in the value of y for a unit change in the corresponding value of x .

ii) $x = 20$

$$= y \quad y = 0.025 + 1.5(20)$$

$$y = 0.025 + 30 \Rightarrow \boxed{y \approx 30}$$

iii) $r=1$ indicates that the relationship between x and y is exactly linear which means that all the points of x and y lie on the straight line.

9) If the two lines of Regression are $4x - 5y + 30 = 0$ and $20x - 9y - 107 = 0$. Which of these Equations is the line of Regression of x on y and y on x

Sol Let us assume that the first Equation is the Regression Equation of x on y and the second Equation is y on x

$$4x - 5y + 30 = 0 \rightarrow (1)$$

$$20x - 9y - 107 = 0 \rightarrow (2)$$

$$(1) \Rightarrow 4x - 5y + 30 = 0$$

$$\Rightarrow 4x - 5y = -30$$

$$\Rightarrow 4x = 5y - 30$$

$$\Rightarrow x = \frac{5}{4}y - \frac{30}{4}$$

We have the Regression co-efficient $b_{xy} = \frac{5}{4}$

$$(2) \Rightarrow 20x - 9y - 107 = 0$$

$$\Rightarrow -9y = -20x - 107$$

$$\Rightarrow y = \frac{20}{9}x - \frac{107}{9}$$

The Regression co-efficient $b_{yx} = \frac{20}{9}$

We have

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

$$r = \sqrt{\left(\frac{5}{4}\right) \left(\frac{20}{9}\right)}$$

$$r = \sqrt{\frac{25}{9}}$$

$$r = \frac{5}{3} \Rightarrow \boxed{r = 1.66}$$

We observe that the correlation co-efficient obtained as $r = 1.66$ with the above assumption is not within the permissible limits $0 \leq r \leq 1$ which is not possible and hence our assumption is wrong.

So the Equation (1) is considered as y on x and Equation (2) as x on y

$$(1) \Rightarrow y \text{ on } x$$

$$\Rightarrow 4x - 5y + 30 = 0$$

$$\Rightarrow 5y = 4x + 30$$

$$y = \frac{4}{5}x + \frac{30}{5}$$

$$b_{yx} = \frac{4}{5}$$

$$(2) \Rightarrow x \text{ on } y$$

$$20x - 9y - 107 = 0$$

$$20x = 9y + 107$$

$$x = \frac{9}{20}y + \frac{107}{20}$$

$$b_{xy} = \frac{9}{20}$$

$$\text{Now } r^2 = b_{xy} b_{yx}$$

$$r^2 = \left(\frac{9}{20}\right)\left(\frac{4}{5}\right)$$

$$r^2 = \frac{9}{25} \Rightarrow r = \sqrt{\frac{9}{25}}$$

$$r = \frac{3}{5}$$

$$r = 0.6$$

**Correlation Analysis Vs Regression Analysis

1. Correlation means the Relationship between two or more variables where the movements (changes) in one variable tend to be accompanied by the corresponding movements in other variable.
2. And on the other hand Regression means Stepping back or returning to the Average value and is a mathematical measure Expressing the relationship between two variables.
3. Correlation co-efficient between the two variables x and y is a measure of the direction and Degree of Linear Relationship between the two variables and is Symmetric $r_{xy} = r_{yx}$ and also it is immaterial which of x and y is dependent variable and which is independent variable.
4. On the other hand Regression Analysis aims at establishing the functional Relationship between the two variables under study and using this relationship to Predict or estimate the value of the dependent variable for any given value of the independent variable. It also reflects upon the Nature of the variable i.e., which is dependent variable and which is independent variable. Hence the Regression co-efficients are Not Symmetric $b_{xy} \neq b_{yx}$.
5. Correlation will not explain cause and effect relationship between the variables under study whereas Regression Analysis will clearly explain the cause and effect relationship between the variables. The variable corresponding to the cause is called INDEPENDENT variable and the variable corresponding to the effect is called DEPENDENT variable.
6. Correlation co-efficient between two variables x and y is a relative measure of the linear Relationship between x and y and it is a pure number which is independent of the units.

7. The Regression co-efficients b_{xy} & b_{yx} are absolute measures representing the change in the value of the variable $y(x)$, for a unit change in the value of other variable $x(y)$.

8. Correlation Analysis is confined only to the study of linear relationship between the variables and therefore has limited applications.

9. The correlation co-efficient is calculated using the following formula

$$r = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{N} \right] \left[\sum y^2 - \frac{(\sum y)^2}{N} \right]}}$$

10. In Simple Regression Analysis there exist two Regression equations as follows

$$Y \text{ on } X \Rightarrow (y - \bar{y}) = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$X \text{ on } Y \Rightarrow (x - \bar{x}) = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

MULTIPLE REGRESSION

* Mbl In Real world, we come across several situations where the Relationship is not that much simple. One variable may be affected by two or more independent variables. for example,

The Sale of a Product may be Related to a number of independent variables such as Price, Income, Advertisement, Season, Size and location of retail outlets, Quality of the Product etc., In such cases, if only one independent variable is considered, the error will be high. The statistical technique of extending the Linear Regression

so as to consider two or more independent variables is known as

MULTIPLE REGRESSION ANALYSIS. The normal equations which

are obtained from the method of least squares in multiple

Regression Analysis are

$$\Sigma y = na + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$\Sigma x_2 y = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

① The following data relates to the radio Advertising Expenditure and the news paper advertising Expenditure and Sales. fit a Regression Equation

Radio Expenditure x_1 (1000's)	4	7	9	12
News paper (1000's) x_2	1	2	5	8
Sales (lacs) y	7	12	17	20

x_1	x_2	y	x_1^2	$x_1 x_2$	x_2^2	$x_1 y$	$x_2 y$
4	1	7	16	4	1	28	7
7	2	12	49	14	4	84	24
9	5	17	81	45	25	153	85
12	8	20	144	96	64	240	160
32	16	56	290	159	94	505	276

The Normal Equations are

$$56 = 4a + 32b_1 + 16b_2 \rightarrow (1)$$

$$505 = 32a + 290b_1 + 159b_2 \rightarrow (2)$$

$$276 = 16a + 159b_1 + 94b_2 \rightarrow (3)$$

Solving the above Simultaneous Equations by considering Eq (1) & (3) and multiplying Eq (1) with (4)

$$\Rightarrow \text{Eq (1)} \times 4$$

$$\Rightarrow 224 = 16a + 128b_1 + 64b_2$$

$$276 = 16a + 159b_1 + 94b_2$$

$$\begin{array}{cccc} (-) & & (-) & (-) & (-) \\ \hline \end{array}$$

$$-52 = -31b_1 - 30b_2$$

$$52 = 31b_1 + 30b_2 \rightarrow \textcircled{4}$$

Solving eq ① and eq ② by multiplying equation ① with 8

$$\Rightarrow \text{eq ①} \times 8$$

$$448 = 32a + 256b_1 + 128b_2$$

$$\begin{array}{r} 505 \\ (-) \end{array} = \begin{array}{r} 32a \\ (-) \end{array} + \begin{array}{r} 290b_1 \\ (-) \end{array} + \begin{array}{r} 159b_2 \\ (-) \end{array}$$

$$-57 = -34b_1 - 31b_2$$

$$57 = 34b_1 + 31b_2 \rightarrow \textcircled{5}$$

Solving equations ④ & ⑤ by multiplying equation ④ with 34 and equation ⑤ with 31

$$\Rightarrow \text{eq ④} \times 34 \quad \& \quad \text{eq ⑤} \times 31$$

$$\Rightarrow 1768 = 1054b_1 + 1020b_2$$

$$\begin{array}{r} 1767 \\ (-) \end{array} = \begin{array}{r} 1054b_1 \\ (-) \end{array} + \begin{array}{r} 961b_2 \\ (-) \end{array}$$

$$1 = 59b_2$$

$$\Rightarrow \boxed{b_2 = 1/59}$$

$$\boxed{b_2 = 0.0169}$$

Substituting $b_2 = 0.0169$ in equation ⑤

$$57 = 34b_1 + 31b_2$$

$$57 = 34b_1 + 31(0.0169)$$

$$57 = 34b_1 + 0.5239$$

$$34b_1 = 56.4761$$

$$\boxed{b_1 = 1.6611}$$

Substituting the values of b_1 and b_2 in equation ①

$$56 = 4a + 32b_1 + 16b_2 \Rightarrow 56 = 4a + 32(1.6611) + 16(0.0169)$$

$$56 = 4a + 53.1552 + 0.2704$$

$$56 = 4a + 53.4256$$

$$4a = 2.5744$$

$$a = 0.6436$$

The Regression Equation is

$$y = a + b_1x_1 + b_2x_2$$

$$y = 0.6436 + 1.6611x_1 + 0.0169x_2$$

From the above equation, it can be understood that the co-efficients b_1 and b_2 give us a clear indication of which variable is more important. b_1 is 1.6611 and b_2 is 0.0169, which means that 1 unit of change in x_1 leads to 1.6611 units of change in y and 1 unit of change in x_2 will lead to 0.0169 units of change in y . Thus it is clear that Radio Advertisement has more impact than Newspaper Advertisement on Sales.

2) The following data relates to the sales, advertisement and the number of selling agents. fit a Regression equation considering the sales as dependent variable (y) and the advertisement (x_1) and the selling agents (x_2) as independent variables.

Sales Lakhs y	100	80	60	120	150	90	70	130
Advertising (000's) x_1	40	30	20	50	60	40	20	60
No. of selling agents x_2	10	10	7	15	20	12	8	14

X_1	X_2	Y	X_1^2	X_2^2	$X_1 X_2$	$X_1 Y$	$X_2 Y$
40	10	100	1600	100	400	4000	1000
30	10	80	900	100	300	2400	800
20	7	60	400	49	140	1200	420
50	15	120	2500	225	750	6000	1800
60	20	150	3600	400	1200	9000	3000
40	12	90	1600	144	480	3600	1080
20	8	70	400	64	160	1400	560
60	14	130	3600	196	840	7800	1820
320	96	800	14600	1278	4270	35400	10480

$$\Sigma y = na + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$\Sigma x_2 y = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

The normal Equations are

$$800 = 8a + b_1 320 + 96 b_2 \rightarrow (1)$$

$$35400 = 320a + 14600 b_1 + 4270 b_2$$

$$35400 = 320a + 14600 b_1 + 4270 b_2$$

$$35400 = 320a + 14600 b_1 + 4270 b_2 \rightarrow (2)$$

$$10480 = 96a + 4270 b_1 + 1278 b_2 \rightarrow (3)$$

Solving the above Simultaneous Equations by considering Eq (1) & (3) multiplying Equation (1) with 12.

$$\Rightarrow \text{Eq (1)} \times 12$$

$$9600 = 96a + 3840 b_1 + 1152 b_2$$

$$\begin{array}{r} 10480 \\ (-) \quad 9600 \\ \hline \end{array} = \begin{array}{r} 96a \\ (-) \quad 96a \\ \hline \end{array} + \begin{array}{r} 4270 b_1 \\ (-) \quad 3840 b_1 \\ \hline \end{array} + \begin{array}{r} 1278 b_2 \\ (-) \quad 1152 b_2 \\ \hline \end{array}$$

$$-880 = -430 b_1 - 126 b_2$$

$$880 = 430 b_1 + 126 b_2 \rightarrow (4)$$

Solving Equation ① & ② multiplying equation ① with 40

$\Rightarrow$ Equation ① $\times 40$

$$\begin{array}{r} 32000 = \cancel{3200a} + 12800b_1 + 3840b_2 \\ 35400 = \cancel{3200a} + 14600b_1 + 4270b_2 \\ \hline (-) \quad (-) \quad (-) \quad (-) \\ -3400 = -1800b_1 - 430b_2 \end{array}$$

$$3400 = 1800b_1 + 430b_2 \rightarrow \textcircled{5}$$

Solving Equations ④ & ⑤ by multiplying equation ④ with 1800 and Equation ⑤ with 430

$$\begin{array}{r} 1584000 = \cancel{774000b_1} + 2268000b_2 \\ 1462000 = \cancel{774000b_1} + 184900b_2 \\ \hline (-) \quad (-) \quad (-) \quad (-) \\ 122000 = 41900b_2 \end{array}$$

$$b_2 = \frac{122000}{41900}$$

$$\boxed{b_2 = 2.9117}$$

Substituting $b_2 = 2.9117$ in Equation ⑤

$$3400 = 1800b_1 + 430(2.9117)$$

$$3400 = 1800b_1 + 1252.0310$$

$$2147.9690 = 1800b_1$$

$$\boxed{b_1 = 1.1933}$$

Substituting b_1 & b_2 values in Equation ①

$$800 = 8a + (1.1933)320 + 96(2.9117)$$

$$800 = 8a + 381.8560 + 279.5232$$

$$800 = 8a + 661.3792$$

$$138.6208 = 8a$$

$$\boxed{a = 17.3276}$$

The Regression Equation is $y = a + b_1x_1 + b_2x_2 \Rightarrow y = 17.3276 + 1.1933x_1 + 2.9117x_2$

Thursday
26/10/17

UNIT-III

TIME SERIES ANALYSIS

The term Time Series indicates a set of observations concerning any activity against different periods of Time. The duration of the time period may be hourly, daily, weekly, Monthly and Yearly.

For Example

1. The Profits earned by a company in the last five years.
2. The Workers employed in a company for the last 15 years.
3. The Number of Students Enrolled for MBA Programme in the last 10 years.
4. The Number of Road Accidents that have taken place in Hyderabad for the last 30 days.

Importance or Significance of Time Series:-

* The Analysis of Time Series helps to understand the past behaviour or Performance. One can know how the data has changed over time and findout probable reasons for such changes and corrective measures can be taken.

* The Time Series Analysis helps directly in business planning. A firm can know the term trend in the sales of a Particular product. The firm can also findout at what rate the sales have been increasing over years which helps to make Projections of sales for the next few years and plan accordingly.

* The Time Series Analysis helps to study the movements as cycles that fluctuate around a Particular trend. A knowledge of cyclical pattern in a series of data will help in making generalization.

in the concerned business or industry.

* The Time Series Analysis helps to make meaningful comparisons in two or more series regarding the rate or the type of growth. For example, the growth in consumption at the National level can be compared with that in National level Income over a certain period of time and such comparisons are of considerable importance for business and industry.

Components of a Time Series:-

1. Secular Trend (T)

2. Seasonal Trend (S)

3. Cyclical Trend (C)

4. Irregular Trend (I)

1. ^{Independent} Secular Trend (T):-

This trend shows that the variable has either increased or decreased over a long period of time. For example, to understand the per capita income changes over the last 20 years.

2. Seasonal Trend (S):-

The second component in the time series is seasonal variation which shows patterns of changes "within a particular year". For example, the demand for woollen garments in winter season and the demand for soft drinks in summer season.

3. Cyclical Trend (C):-

The third component in time series analysis is cyclical variation where there exist peaks and slacks. For example, in business there will be certain periods where the business will be at its peak, while in some other periods the business will slowdown. It is just like a wave.

4. Irregular Trend (I) :-

This is the fourth component and last component in the Time Series. These variations are irregular and unpredictable. For example, the occurrence earthquakes.

The effect of the above-four components it may be additive or multiplicative i.e. $y = T \times S \times C \times I$ or $y = T + S + C + I$, where y = The result of the four components.

T = The Secular Trend.

S = Seasonal Trend.

C = cyclical trend.

I = Irregular Trend.

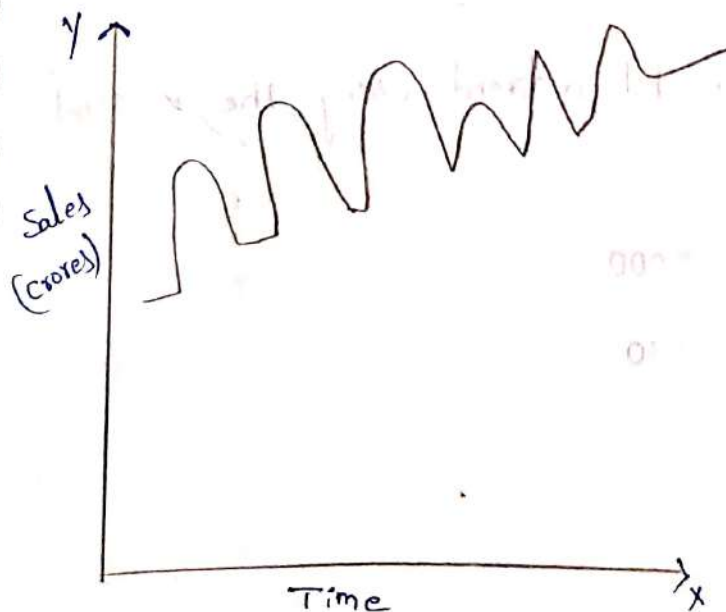
Saturday
28/10/17

The Trends:-

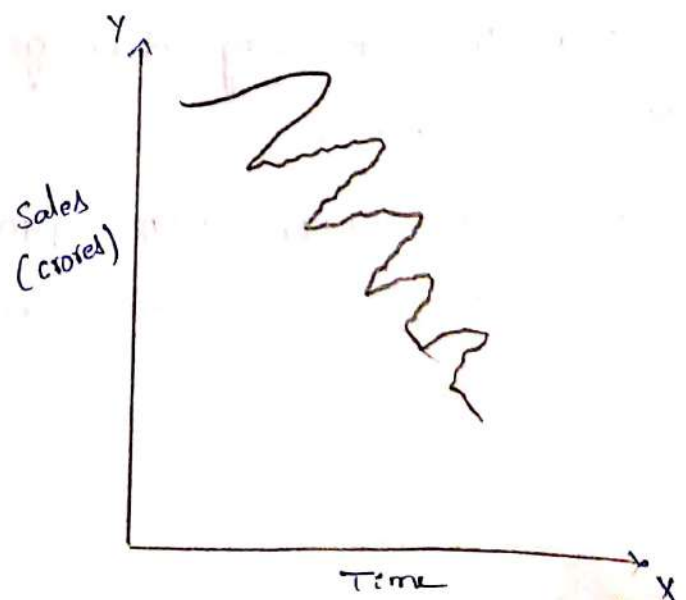
The Trend is the longterm movement of a time series.

In other words, an increase or decrease in the value of a variable occurring over a period of several years gives a trend. If the values of a variable remain stationary over several years, we can say there is no trend.

The following graph shows the trend in sales over a period of time that is increasing and decreasing trend.



Increasing Trend



Decreasing Trend.

The following are the methods used for fitting a straight line to a Time Series.

1. Free Hand Method.
2. Method of Semi Averages.
3. Method of moving Averages.
4. Method of least Squares.

Method of least Squares:-

Out of all the methods available for fitting a straight line to a Series of data, the most frequently used method is the method of least squares. According to this method, the equation of a straight line is given by $y = a + bx$

Where y = Time Period

x = The value of the item measured against time.

a, b = constants.

In order to find out the values of a and b , the following two equations are used

$$\sum y = na + b \sum x$$

$$\sum xy = a \sum x + b \sum x^2$$

1) From the following data given, fit a trend using the method of least squares

Year	1996	1997	1998	1999	2000
y	270	285	295	315	330

Year	y	x	xy	x ²
1996	270	-2	-540	4
1997	285	-1	-285	1
1998	295	0	0	0
1999	315	1	315	1
2000	330	2	660	4
	1495	0	150	10

$x = 1998 - \text{previous number}$

ie = $1998 - 1997$

$$\Sigma y = na + b \Sigma x$$

$$1495 = 5a + 0$$

$$\Rightarrow a = \frac{1495}{5}$$

$$a = 299$$

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

$$150 = 299(0) + b(10)$$

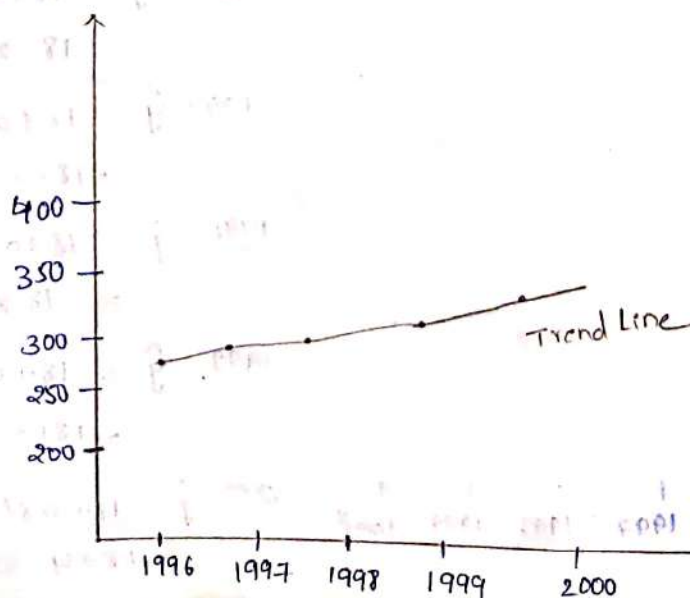
$$b = \frac{150}{10} \Rightarrow b = 15$$

Substituting these values of a and b in the normal equation

we have

$$y = a + bx$$

$$\Rightarrow y = 299 + 15x$$



$$\begin{aligned} \underline{1996} \quad \hat{y} &= 299 + 15(-2) \\ &= 299 - 30 = 269 \end{aligned}$$

$$\begin{aligned} \underline{1997} \quad \hat{y} &= 299 + 15(-1) \\ &= 299 - 15 = 284 \end{aligned}$$

$$\begin{aligned} \underline{1998} \quad \hat{y} &= 299 + 15(0) \\ &= 299 \end{aligned}$$

$$\begin{aligned} \underline{1999} \quad \hat{y} &= 299 + 15(1) \\ &= 299 + 15 = 314 \end{aligned}$$

$$\begin{aligned} \underline{2000} \quad \hat{y} &= 299 + 15(2) \\ &= 299 + 30 = 329 \end{aligned}$$

2) For the following data given fit a trend using the method of least squares

Year(x)	1995	1996	1997	1998	1999	2000
y	15	14	18	20	17	24

Sol

year	y	x	xy	x ²
1995	15	-5	-75	25
1996	14	-3	-42	9
1997	18	-1	-18	1
1998	20	1	20	1
1999	17	3	51	9
2000	24	5	120	25
	<u>108</u>	<u>0</u>	<u>56</u>	<u>70</u>

1995-2000
2000-1995

$$\Sigma y = na + b \Sigma x$$

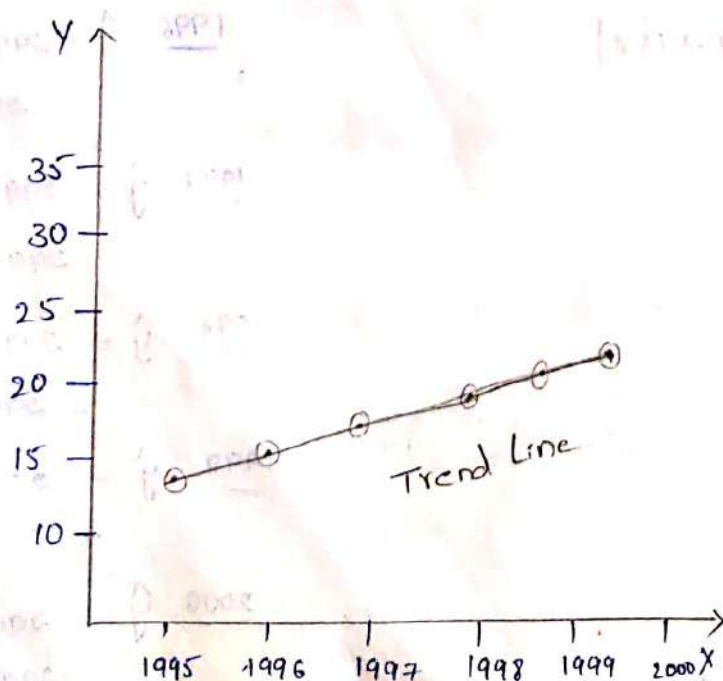
$$108 = 6a + b(0) \Rightarrow 6a = 108$$

$$a = 18$$

$$\Sigma xy = a \Sigma x + b \Sigma x^2 \Rightarrow 56 = 0 + 70b$$

$$b = 0.8$$

$$y = a + bx \Rightarrow y = 18 + 0.8x$$



$$\underline{1995} \quad \hat{y} = 18 + 0.8(-5) = 18 - 4 = 14$$

$$\underline{1996} \quad \hat{y} = 18 + 0.8(-3) = 18 - 2.4 = 15.6$$

$$\underline{1997} \quad \hat{y} = 18 + 0.8(-1) = 18 - 0.8 = 17.2$$

$$\underline{1998} \quad \hat{y} = 18 + 0.8(1) = 18.8$$

$$\underline{1999} \quad \hat{y} = 18 + 0.8(3) = 18 + 2.4 = 20.4$$

$$\underline{2000} \quad \hat{y} = 18 + 0.8(5) = 18 + 4 = 22$$

3) Fit a Straight line Trend by the method of Least squares

Year	1991	1992	1993	1994	1995	1996
Production tonnes	24	25	29	26	22	24

Sol

Year	y	x	xy	x ²
1991	24	-5	-120	25
1992	25	-3	-75	9
1993	29	-1	-29	1
1994	26	1	26	1
1995	22	3	66	9
1996	24	5	120	25
	150	0	-12	70

$$\Rightarrow \Sigma y = na + b \Sigma x$$

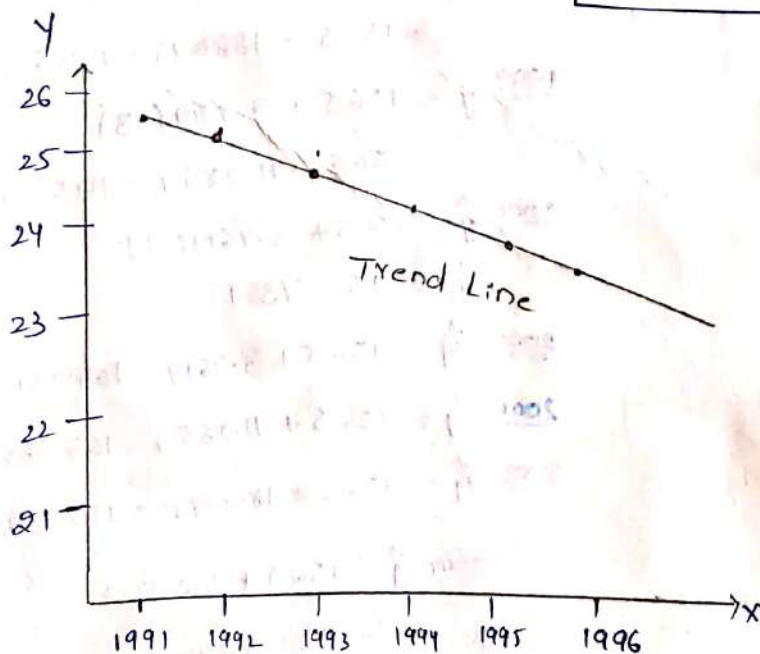
$$150 = 6a + 0 = \therefore \boxed{a = 25}$$

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

$$-12 = 25(0) + b(70) \Rightarrow \boxed{b = -0.1714}$$

$$y = a + bx = \therefore \boxed{y = 25 - 0.1714x}$$

$$\begin{aligned} -12 &= 0 + 70b \\ 70b &= -12 \\ b &= \frac{-12}{70} \end{aligned}$$



$$\begin{aligned} \underline{1991} \quad \hat{y} &= 25 - 0.1714(-5) \\ &= 25 + 0.8571 \\ &= 25.8571 \end{aligned}$$

$$\begin{aligned} \underline{1992} \quad \hat{y} &= 25 - (0.1714)(-3) \\ &= 25 + 0.5142 = \therefore 25.5142 \end{aligned}$$

$$\begin{aligned} \underline{1993} \quad \hat{y} &= 25 - (0.1714)(-1) \\ &= 25 + 0.1714 = \therefore 25.1714 \end{aligned}$$

$$\begin{aligned} \underline{1994} \quad \hat{y} &= 25 - (0.1714)(1) \\ &= 24.8286 \end{aligned}$$

$$\begin{aligned} \underline{1995} \quad \hat{y} &= 25 - (0.1714)(3) \\ &= 25 - 0.5142 = 24.4858 \end{aligned}$$

$$\underline{1996} \quad \hat{y} = 25 - (0.1714)(5) = 25 - 0.8571 = 24.1429$$

Monday 30/10/17
fit a straight line trend by the method of least squares to the following data. Also calculate the trend from 1997 to 2004. Also forecast the sales for 2008.

Year	1997	1998	1999	2000	2001	2002	2003	2004
Sales (lacks)	76	80	330	144	138	120	174	190

Sol

Year	y	x	xy	x ²
1997	76	-7	-532	49
1998	80	-5	-400	25
1999	330	-3	-990	9
2000	144	-1	-144	1
2001	138	1	138	1
2002	120	3	360	9
2003	174	5	870	25
2004	190	7	1330	49
	1252	0	632	168

$$\Sigma y = na + b \Sigma x$$

$$1252 = 8a + b(0)$$

$$a = \frac{1252}{8} \Rightarrow \boxed{a = 156.5}$$

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

$$632 = 8(0) + b(168)$$

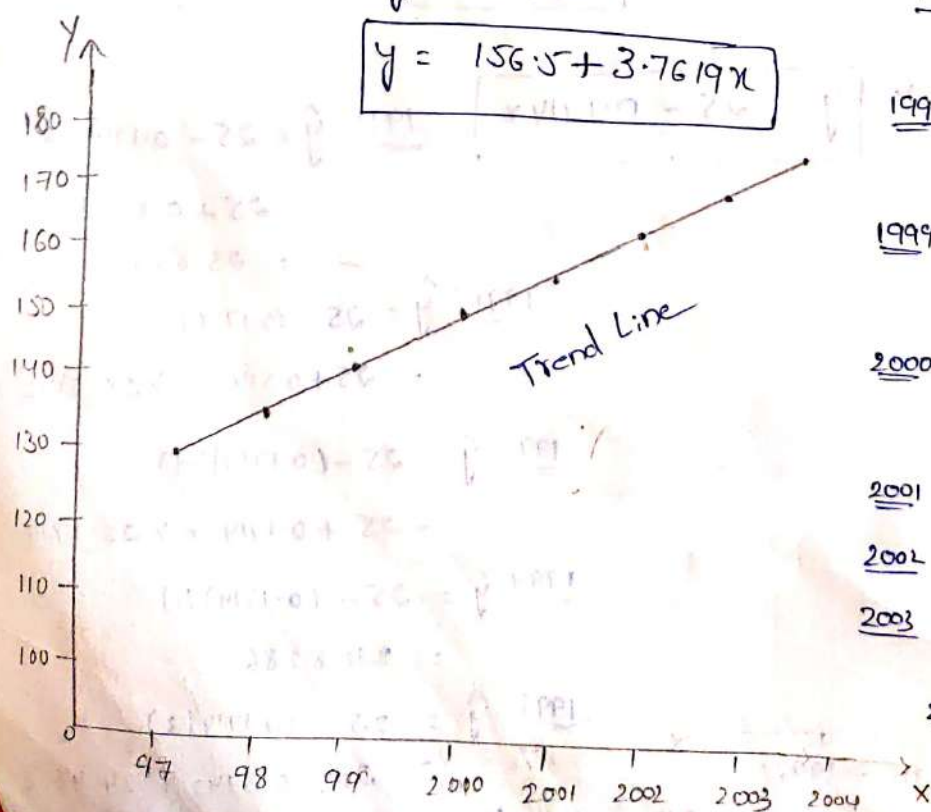
$$168b = 632$$

$$b = \frac{632}{168} \Rightarrow \boxed{b = 3.7619}$$

Substituting the values of a and b in the normal equation we have

$$y = a + bx$$

$$\boxed{y = 156.5 + 3.7619x}$$



$$\underline{1997} \hat{y} = 156.5 + 3.7619(-7)$$

$$= 156.5 - 26.3333 = 130.1667$$

$$\underline{1998} \hat{y} = 156.5 + 3.7619(-5)$$

$$= 156.5 - 18.8095 = 137.6905$$

$$\underline{1999} \hat{y} = 156.5 + 3.7619(-3)$$

$$= 156.5 - 11.2857 = 145.2143$$

$$\underline{2000} \hat{y} = 156.5 + 3.7619(-1)$$

$$= 152.7381$$

$$\underline{2001} \hat{y} = 156.5 + 3.7619 = 160.2619$$

$$\underline{2002} \hat{y} = 156.5 + 11.2857 = 167.7857$$

$$\underline{2003} \hat{y} = 156.5 + 18.8095 = 175.3095$$

$$\underline{2004} \hat{y} = 156.5 + 26.3333$$

$$= 182.8333$$

The sales for 2008 is

$$x=15$$

substituting $x=15$ in the equation

$$y = 156.5 + 3.7619x$$

$$y = 156.5 + 3.7619(15)$$

$$y = 212.9285$$

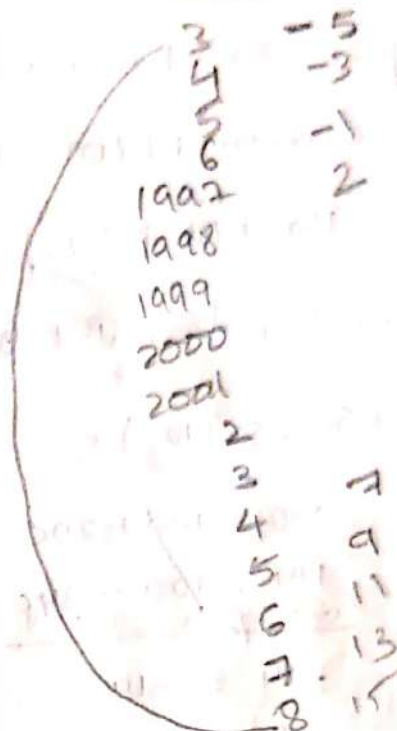
$$y \approx 213$$

Non-Linear Trend:-

$$\hat{y} = a + bx + cx^2$$

Normal Trend Equations are

$$\begin{aligned} \sum y &= na + b\sum x + c\sum x^2 \\ \sum xy &= a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2y &= a\sum x^2 + b\sum x^3 + c\sum x^4 \end{aligned}$$



① For the following data fit a trend

Year	1995	1996	1997	1998	1999
Production	2	4	8	14	22

Sol

Year	y	x	x ²	x ³	x ⁴	xy	x ² y
1995	2	-2	4	-8	16	-4	8
1996	4	-1	1	-1	1	-4	4
1997	8	0	0	0	0	0	0
1998	14	1	1	1	1	14	14
1999	22	2	4	8	16	44	88
	50	0	10	0	34	50	114

Normal Trends

$$\sum y = na + b\sum x + c\sum x^2$$

$$50 = 5a + b(0) + c(10) = y \quad \boxed{50 = 5a + 10c} \rightarrow \textcircled{1}$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$

$$50 = a(10) + b(10) + c(10)$$

$$50 = 10a + 10b + 10c \quad \text{--- (1)} \quad \Rightarrow \quad b = 5$$

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

$$114 = a(10) + b(10) + c(34)$$

$$114 = 10a + 10b + 34c \quad \text{--- (2)}$$

Now solving equations (1) & (2) by multiplying equation (1) with '2'

$$(50 = 5a + 10c) \times 2$$

$$100 = 10a + 20c$$

$$114 = 10a + 34c$$

$$-14 = -14c$$

$$c = 1$$

Substituting $c = 1$ in equation (1)

$$50 = 5a + 10(1)$$

$$50 = 5a + 10$$

$$5a = 40 \Rightarrow a = 8$$

Substituting the values of a, b and c in the Non-Linear Trend Equation we have

$$\hat{y} = a + bx + cx^2$$

$$\hat{y} = 8 + 5x + x^2$$

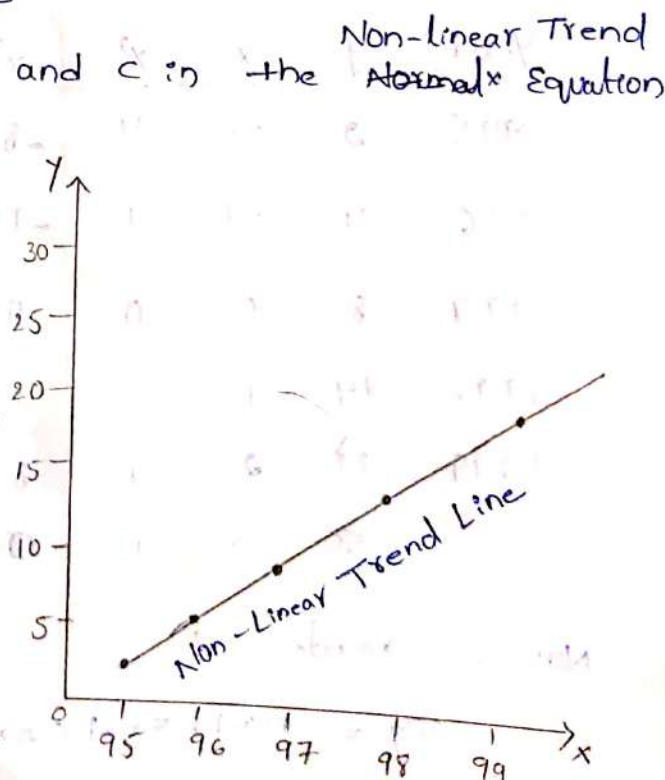
$$\begin{array}{l} \text{1995} \\ \hat{y} = 8 + 5(-2) + 4 = 8 - 10 + 4 = 2 \end{array}$$

$$\begin{array}{l} \text{1996} \\ \hat{y} = 8 + 5(-1) + 1 = 8 - 5 + 1 = 4 \end{array}$$

$$\begin{array}{l} \text{1997} \\ \hat{y} = 8 + 5(0) + 0 = 8 + 0 + 0 = 8 \end{array}$$

$$\begin{array}{l} \text{1998} \\ \hat{y} = 8 + 5(1) + 1 = 8 + 5 + 1 = 14 \end{array}$$

$$\begin{array}{l} \text{1999} \\ \hat{y} = 8 + 5(2) + 4 = 8 + 10 + 4 = 22 \end{array}$$



Tuesday
31/10/17

1) The sales of a company (lacks of Rupees) for the 7 years are given below

year	1990	1991	1992	1993	1994	1995	1996
sales	32	47	65	88	132	190	275

find out the trend values by using the equation $y = ab^x$

Year x	Sales y	Time centered log x	Time centered log y	x^2	$x \log y$
1990	32	-3	1.5051	9	-4.5153
1991	47	-2	1.6721	4	-3.3442
1992	65	-1	1.8129	1	-1.8129
1993	88	0	1.9445	0	0
1994	132	1	2.1206	1	2.1206
1995	190	2	2.2788	4	4.5576
1996	275	3	2.4393	9	7.3179
	829	0	13.7686	28	4.3237

$$\log a = \frac{\sum \log y}{n} \Rightarrow \log a = \frac{13.7686}{7} = 1.9669$$

$$\log b = \frac{\sum x \log y}{\sum x^2} \Rightarrow \log b = \frac{4.3237}{28} = 0.1544$$

Now the Trend equation is obtained as follows

Trends of logs

$$y = ab^x$$

$$\Rightarrow y = \text{Anti log} (1.9669) (\text{Anti log } 0.1544)^x$$

$$\Rightarrow y = (92.6616) (1.4269)^x$$

$$\Rightarrow y / F$$

2) fit a simple Exponential Trend to the following data and estimate the Sales for 1989

Year	1985	1986	1987	1988	1989
Sales (crores)	100	105	112	120	130

Sol

Year	Sales(y)	log y	Time centered X	X ²	X log y
1985	100	2	-2	4	-4
1986	105	2.0212	-1	1	-2.0212
1987	112	2.0492	0	0	0
1988	120	2.0792	1	1	2.0792
1989	130	2.1139	2	4	4.2278
	567	10.2635	0	10	0.2858

$$\log a = \frac{\sum \log y}{n} = \frac{10.2635}{5} = 2.0527$$

$$\log b = \frac{\sum X \log y}{\sum X^2} = \frac{0.2858}{10} = 0.0286$$

Now the trend equation is obtained as follows

Trends of logs

$$y = ab^x$$

$$\Rightarrow y = \text{Antilog}(2.0527)(\text{Antilog } 0.0286)^x$$

$$\Rightarrow y = (112.9016)(1.0681)^x$$

Sales in 19889

$$y = (112.9016)(1.0681)^2$$

$$= (112.9016)(1.1408)$$

$$= 128.8024$$

$$y = 129$$

Wednesday

11/11/17

INDEX NUMBERS

- * Index Numbers are indicators which reflect the relative changes in the level of a certain phenomena/data/variable in any given period or over a specified period of time called "Current Period" with respect to its value in some fixed period called "Base Period" selected for comparison.
- * The phenomena (or) variable under consideration may be the prices of particular commodities like steel, gold, milk, cosmetics, factory production, imports and exports, national income, wage structure of workers, cost of living, bank deposits etc.
- * Definition 1:- Index Numbers are statistical devices designed to measure the relative change of a phenomenon (variable or group of variables) with respect to time. In other words Index Numbers are specialized numbers, ratios, percentages which give the general level of the magnitude of a group of distinct but related variables in two or more situations.
- * For example, if the general change in the price level of consumer goods is studied in a particular section of society say low income group or middle income group.
- * ^{Def} Index Numbers is a statistical device which helps to arrive at a single representative figure which gives the general level of the prices of commodities.
- * Definition 2:- Index Numbers is a statistical device for indicating the relative movements of the data where measurements of actual movements is difficult to measure.
- * The following are the uses of index numbers.
 - a) Index numbers are economic barometers.
 - b) Index numbers are used in studying trends.
 - c) Index numbers help in decision making.

dy Index Numbers are used in calculating Purchasing Power.

Terminology:-

1. BASE YEAR - which is denoted by suffix 0 (P_0, Q_0)
2. CURRENT YEAR - which is denoted by suffix 1 (P_1, Q_1)
3. P_0 = Price of the commodity in the base year.
4. P_1 = The Price of the commodity in the current year.
5. Q_0 = Quantity of a commodity consumed or Purchased in the base year.
6. Q_1 = Quantity of a commodity consumed or Purchased during the current year.
7. W = The weight assigned to a commodity, According to its relative importance in a group.
8. I = Simple Index number (or) Price relative. obtained by Expressing current year price as a Percentage of base year price.

$$I = \frac{P_1}{P_0} \times 100$$
9. P_{01} = Price Index Number for the current year with respect to base year.
10. Q_{01} = Quantity Index Number for the current year with respect to base year.
11. P_{10} = Price Index number for the base year with respect to current year.
12. Q_{10} = Quantity Index number for the base year with respect to current year.
13. V_{01} = Volume Index^{number} for the current year with respect to the base year.
14. V_{10} = Volume Index number for the base year with respect to the current year.

Friday
3/11/17

1) Calculate the Price Index Number of a commodity for the period 1995 to 2000

Year	1995	1996	1997	1998	1999	2000
Price (Rs)	3000	3200	3250	3100	3290	3600

Sol For example, to calculate the Index Number for the year 1998 taking 1995 as the base year, then the Price Index Number is

$$= \frac{3100}{3000} \times 100$$

$$= 103.33.$$

To calculate the Index Number for the year 2000 with 1996 as the base year

$$\text{Price Index Number} = \frac{3600}{3200} \times 100$$

$$= 112.5$$

Composite Index Numbers:-

These Index Numbers are the Index Numbers which refer to the Price changes in several commodities. For example, calculate the Price Index Number for year one and year two for the following three commodities

Commodity	Price year one	Price year Two
Bread	10	14
Milk	15	20
Eggs	10	50
	<u>35</u>	<u>84</u>

$$\text{Price Index Number for year one} = \frac{35}{35} \times 100 = 100$$

$$\text{Price Index Number for year Two} = \frac{84}{35} \times 100 = 240$$

Laspeyres formula:-

The Index Number which is based on the base year weights is known as Laspeyres Index. which is given by the following formula

$$\text{Laspeyres Index} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

$$Q_{01}^{LA} = \frac{\sum Q_1 P_0}{\sum Q_0 P_0} \times 100$$

Paasche formula:-

P_{01}^{Pa}

$$\text{Paasche formula} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

Quantity Index

$$Q_{01}^{Pa} = \frac{\sum Q_1 P_1}{\sum Q_0 P_1} \times 100$$

Bowley's Index:-

$$\Rightarrow \frac{\frac{L+P}{2}}{\frac{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} + \frac{\sum P_1 Q_1}{\sum P_0 Q_1}}{2}}$$

Fishers Ideal Index:-

$$\text{Quantity Index} \Rightarrow Q_{01}^F = \sqrt{Q_{01}^{La} \times Q_{01}^{Pa}}$$

$$\text{Fishers Ideal Index} = \sqrt{L \times P} = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}}$$

NOTE:-

The fishers Ideal Index Number is the square root of Laspeyres and Paasche Indices. This fishers Index is known as Ideal Index when it satisfies the following two tests

1. Time Reversal Test.
2. Factor Reversal Test.

Time Reversal Test:-

if it gives

$$P_{01} \times P_{10} = 1$$

A Method satisfying the time Reversal Test

Where P_{01} is the Price Index Number for the current year '1' with respect to the base year '0' and P_{10} is the Price Index Number for the Base year Taking the current year as the Base year

Factor Reversal Test:-

A method that satisfies factor Reversal Test

gives

$$P_{01} \times Q_{01} = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

Where P_{01} is the Price Index for

the current year and Q_{01} is the Quantity Index for the current year.

1) Given below is the data related to the Price and Quantity for two years related to three commodities. Calculate the Paasche Index Number

commodity	P_0	Q_0	P_1	Q_1	$P_1 Q_1$	$P_0 Q_1$
A	10	40	15	60	900	600
B	15	80	20	100	2000	1500
C	20	20	25	40	1000	800
					<u>3900</u>	<u>2900</u>

$$\begin{aligned} \text{sol} \quad \text{Paasche Index} &= \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100 \\ &= \frac{3900}{2900} \times 100 \\ &= 134.48 \end{aligned}$$

Saturday
4/11/17

2) The Price and Quantity data for two years for five commodities are given below. Determine the Laspeyres, Paasche Indices and also Fisher's Ideal Index and test for Time Reversal and Factor Reversal Test

commodity	P_0	Q_0	P_1	Q_1	$P_0 Q_0$	$P_1 Q_0$	$P_1 Q_1$	$P_0 Q_1$
A	10	5	15	5	50	75	75	50
B	5	10	5	12	50	50	60	60
C	8	4	10	5	32	40	50	40
D	12	5	15	5	60	75	75	60
E	6	15	12	10	90	180	120	60
					<u>282</u>	<u>420</u>	<u>380</u>	<u>270</u>

$$\text{Laspeyres} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} = \frac{420}{282} \times 100 = 148.93$$

$$\text{Paasche Index} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} = \frac{380}{270} \times 100 = 140.74$$

$$\begin{aligned} \text{Fisher's Index} &= \sqrt{L \times P} = \sqrt{(148.93)(140.74)} = \sqrt{20960.4082} \\ &= 144.77 \end{aligned}$$

③ calculate the price Index number for the year 1996 with 1986 as the base year for the following data using
 i) Laspeyres Index
 ii) Paasche Index
 iii) Fisher's Index.

commodity	unit	Price	value
A	kg	10	1500
B	kg	12	1080
C	metre	15	900
D	packets	9	450

1996		
commodity	Quantity	Value (Rs)
	160	1760
	100	1300
	60	960
	40	480

Sol	commodity	P_0	Q_0	P_1	Q_1	$P_0 Q_0$	$P_1 Q_1$	$P_0 Q_1$	$P_1 Q_0$
	A	10	$\frac{1500}{10} = 150$	$\frac{1760}{160} = 11$	160	1500	1760	1600	1650
	B	12	$\frac{1080}{12} = 90$	$\frac{1300}{100} = 13$	100	1080	1300	1200	1170
	C	15	$\frac{900}{15} = 60$	$\frac{960}{60} = 16$	60	900	960	900	960
	D	9	$\frac{450}{9} = 50$	$\frac{480}{40} = 12$	40	450	480	360	600
						<u>3930</u>	<u>4500</u>	<u>4060</u>	<u>4380</u>

i) Laspeyres Index = $\frac{\sum P_1 Q_0}{\sum P_0 Q_0} = \frac{4380}{3930} \times 100 = 111.4504$

ii) Paasche Index = $\frac{\sum P_1 Q_1}{\sum P_0 Q_1} = \frac{4500}{4060} \times 100 = 110.8374$

iii) Fisher's Index = $\sqrt{L \times P}$

$$= \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} = \sqrt{(111.4504)(110.8374)}$$

$$= \sqrt{12352.8768}$$

$$= 111.1435$$

4) The Price and quantity data for the year 2005 & 2006 are given below. Construct the index numbers using Laspeyres, Paasche Index, Fisher's, Bowley's and Marshall Edge worth.

Commodity	2005		2006	
	Price	Quantity	Price	Quantity
A	3	5	6	7
B	4	6	8	8
C	6	5	9	10
D	2	10	8	12
E	5	15	10	20

Sol

Commodity	P ₀	q ₀	P ₁	q ₁	P ₀ q ₀	P ₁ q ₁	P ₀ q ₁	P ₁ q ₀
A	3	5	6	7	15	42	21	30
B	4	6	8	8	24	64	32	48
C	6	5	9	10	30	90	60	45
D	2	10	8	12	20	96	24	80
E	5	15	10	20	75	200	100	150
					164	492	237	353

$$\text{Laspeyres Index} = \frac{\sum P_1 q_0}{\sum P_0 q_0} = \frac{353}{164} \times 100 = 215.2439$$

$$\text{Paasche Index} = \frac{\sum P_1 q_1}{\sum P_0 q_1} = \frac{492}{237} \times 100 = 207.5949$$

$$\begin{aligned} \text{Fisher's Index} &= \sqrt{L \times P} \\ &= \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}} \\ &= \sqrt{(215.2439) (207.5949)} \\ &= \sqrt{44683.5359} \\ &= 211.3848 \end{aligned}$$

$$\begin{aligned}
 \text{Bowley's} &= \frac{L+P}{2} \\
 &= \frac{\frac{\sum P_1 V_0}{\sum P_0 V_0} + \frac{\sum P_1 V_1}{\sum P_0 V_1}}{2} \\
 &= \frac{215.2439 + 207.5949}{2} \\
 &= \frac{422.8388}{2} \\
 &= 211.4194
 \end{aligned}$$

$$\text{Marshall Edge Worth Price Index} \Rightarrow P_{01}^{ME} = \left[\frac{\sum P_1 V_0 + \sum P_1 V_1}{\sum P_0 V_0 + \sum P_0 V_1} \right] \times 100$$

$$\begin{aligned}
 \text{Marshall Edge Worth Index} &= \left[\frac{\sum P_1 V_0 + \sum P_1 V_1}{\sum P_0 V_0 + \sum P_0 V_1} \right] \times 100 \\
 &= \left[\frac{353 + 492}{164 + 237} \right] \times 100 \\
 &= \frac{845}{401} \times 100 \\
 &= 210.7232.
 \end{aligned}$$

Time Reversal Test:-

$$P_{01} \times P_{10} = 1 \quad (\text{omitting factor 100})$$

where

P_{01} = Price Index (without factor 100) for year '1' with year '0' as base.

P_{10} = Price Index (without factor 100) for year '0' with year '1' as base

$$P_{01}^F \times P_{10}^F = 1$$

where

$$P_{01}^F = \sqrt{L \times P} = \sqrt{\frac{\sum P_1 V_0 \times \sum P_1 V_1}{\sum P_0 V_0 \times \sum P_0 V_1}} \times 100, \quad P_{10}^F = \sqrt{\frac{\sum P_0 V_1 \times \sum P_0 V_0}{\sum P_1 V_1 \times \sum P_1 V_0}} \times 100$$

Factor Reversal Test

Without factor 100

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} = \frac{\sum V_1}{\sum V_0}$$

where

$$P_{01}^F = \sqrt{\frac{\sum P_1 Q_0 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_0 Q_1}} \times 100$$

$$Q_{01}^F = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_1 Q_0}} \times 100 = \sqrt{\frac{\sum Q_1 P_0 \times \sum Q_1 P_1}{\sum Q_0 P_0 \times \sum Q_0 P_1}}$$

① for the following data prove that the Fisher's Ideal Index satisfies both the time reversal Test and the factor Reversal Test and calculate its value.

Commodity	Base year		current year	
	Price	Quantity	Price	Quantity
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24

Sol

Commodity	P_0	Q_0	P_1	Q_1	$P_0 Q_0$	$P_1 Q_1$	$P_0 Q_1$	$P_1 Q_0$
A	6	50	10	56	300	560	336	500
B	2	100	2	120	200	240	240	200
C	4	60	6	60	240	360	240	360
D	10	30	12	24	300	288	240	360
					<u>1040</u>	<u>1448</u>	<u>1056</u>	<u>1420</u>

$$\text{Fisher's Price Index Number} = \sqrt{\frac{\sum P_1 Q_0 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_0 Q_1}}$$

$$P_{01}^F = \sqrt{\frac{1420 \times 1448}{1040 \times 1056}} = \sqrt{\frac{2056160}{1098240}} = \sqrt{1.8722} = 1.3683 \times 100 \Rightarrow 136.83$$

Time Reversal Test:-

$$P_{01}^F \times P_{10}^F = 1$$

$$P_{10}^F = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_0 Q_0}{\sum P_1 Q_1 \times \sum P_1 Q_0}} \times 100$$

without factor 100

$$= \sqrt{\frac{1056 \times 1040}{1448 \times 1420}}$$

$$= \sqrt{\frac{1098240}{2056160}}$$

$$= \sqrt{0.5341}$$

$$= 0.7308$$

$$\text{Now } P_{01}^F \times P_{10}^F$$

$$= 1.3683 \times 0.7308$$

$$= 1$$

Factor Reversal Test:-

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

$$Q_{01}^F = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_1 Q_0}}$$

$$= \sqrt{\frac{1056 \times 1448}{1040 \times 1420}}$$

$$= \sqrt{\frac{1529088}{1476800}}$$

$$= \sqrt{1.0354}$$

$$= 1.0175$$

$$P_{01}^F \times Q_{01}^F = 1.3683 \times 1.0175$$

$$= 1.3923$$

$$\text{Now } \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

$$= \frac{1448}{1040}$$

$$= 1.3923$$

Now

$$P_0^F \times Q_0^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} \text{ proved.}$$

Hence the Fisher's Index Number proved both Time reversal Test and factor Reversal Test.

② calculate Laspeyres, Paasche and Fisher's Indices for the following data and also examine for the Time Reversal and the factor Reversal Test.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	6.5	500	10.8	560
B	2.8	124	2.9	148
C	4.7	69	8.2	78
D	10.9	38	13.4	24
E	8.6	49	10.8	27

Sol

Commodity	P_0	Q_0	P_1	Q_1	$\underline{P_0 Q_0}$	$\underline{P_1 Q_1}$	$\underline{P_0 Q_1}$	$\underline{P_1 Q_0}$
A	6.5	500	10.8	560	3250	6048	3640	5400
B	2.8	124	2.9	148	347.20	429.20	414.40	359.60
C	4.7	69	8.2	78	324.30	639.60	366.60	565.80
D	10.9	38	13.4	24	414.20	321.60	261.60	509.20
E	8.6	49	10.8	27	421.40	291.60	232.20	529.20
					<u>4757.10</u>	<u>7730</u>	<u>4914.80</u>	<u>529.20</u>
								<u>7363.80</u>

$$\text{Laspeyres Index} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} = \frac{7363.80}{4757.10} = 1.5480 \times 100 = 154.7960$$

$$\text{Paasche Index} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} = \frac{7730}{4914.80} = 1.5728 \times 100 = 157.28$$

$$\begin{aligned} \text{Fisher's Index} &= \sqrt{L \times P} = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \\ &= \sqrt{154.80 \times 157.28} \\ &= \sqrt{24346.9440} \\ &= 156.0351 \end{aligned}$$

Time Reversal Tests:-

$$P_{01} \times P_{10} = 1$$

$$P_{01} = \sqrt{\frac{\sum P_1 Q_0 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_0 Q_1}} \times 100$$

$$P_{10} = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_0 Q_0}{\sum P_1 Q_1 \times \sum P_1 Q_0}} \times 100$$

$$P_{10} = \sqrt{\frac{4914.80 \times 4757.10}{7730 \times 7363.80}}$$

$$= \sqrt{\frac{23380195.08}{56922174}}$$

$$= \sqrt{0.4107}$$

$$= 0.6409$$

Now

$$P_{01}^F \times P_{10}^F = 1.560351 \times 0.6409$$

$$= 1$$

$$\boxed{P_{01}^F \times P_{10}^F = 1}$$

Hence Proved.

Factor Reversal Test:-

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} = \frac{\sum V_1}{\sum V_0}$$

$$Q_{01}^F = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_1 Q_0}} \times 100$$

$$= \sqrt{\frac{4914.80 \times 7730}{4757.10 \times 7363.80}}$$

$$= \sqrt{\frac{37991404}{35030332.98}}$$

$$= \sqrt{1.0845}$$

$$= 1.0414$$

Now $\frac{\sum P_i Q_i}{\sum P_0 Q_0}$

$$P_{01}^F \times Q_{01}^F = 1.560351 \times 1.0414$$

$$= 1.6250$$

Now

$$\frac{\sum P_i Q_i}{\sum P_0 Q_0}$$

$$= \frac{7730}{4757.10}$$

$$= 1.6250$$

$$\frac{01.5284 \times 08.4114}{08.2285 \times 08.4114}$$

Now

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_i Q_i}{\sum P_0 Q_0} \text{ is proved.}$$

Hence Fisher's Index Number proved both Time Reversal Test and Factor Reversal Test.

3) Compute the price Index and Quantity Index Numbers for the years 2000 with 1995 as base year using i) Laspeyres method ii) Paasche's method iii) also compute Fisher's price Index and Quantity Index Numbers

Commodity	Quantity units		Value	
	1995	2000	1995	2000
A	100	150	500	900
B	80	100	320	500
C	60	72	150	360
D	30	33	360	297

sol

$$\text{Value} = \text{Price} \times \text{Quantity}$$

$$\Rightarrow \text{Price} = \frac{\text{Value}}{\text{Quantity}}$$

q_0	q_1	P_0	P_1	$P_0 q_0$	$P_1 q_1$	$P_1 q_0$	$P_0 q_1$
100	150	$\frac{500}{100} = 5$	$\frac{900}{150} = 6$	500	900	600	750
80	100	$\frac{320}{80} = 4$	$\frac{500}{100} = 5$	320	500	400	400
60	72	$\frac{150}{60} = 2.5$	$\frac{360}{72} = 5$	150	360	300	180
30	33	$\frac{360}{30} = 12$	$\frac{297}{33} = 9$	360	297	270	396
				1330	2057	1570	1726

$$\text{Laspeyres's Index} = \frac{\sum P_1 q_0}{\sum P_0 q_0} = \frac{1570}{1330} = 1.1805 \times 100$$

$$\text{Paasche's Index} = \frac{\sum P_1 q_1}{\sum P_0 q_1} = \frac{2057}{1726} = 1.1918 \times 100$$

$$= 119.18$$

$$\text{Fisher's Index} = \sqrt{\text{LXP}} = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}}$$

$$= \sqrt{(118.05)(119.18)} = \sqrt{14069.1990} = 118.6137$$

Fisher's Quantity Index:-

$$Q_{01}^F = \sqrt{Q_{01}^{La} \times Q_{01}^{Pa}}$$

$$Q_{01}^{La} = \frac{\sum P_0 Q_1}{\sum Q_0 P_0} \times 100 \Rightarrow \frac{1726}{1330} \times 100 = 129.7744$$

$$Q_{01}^{Pa} = \frac{\sum Q_1 P_1}{\sum Q_0 P_1} \times 100 \Rightarrow \frac{2057}{1570} \times 100 = 131.02$$

$$\begin{aligned} Q_{01}^F &= \sqrt{Q_{01}^{La} \times Q_{01}^{Pa}} \\ &= \sqrt{(129.77)(131.02)} \\ &= \sqrt{17002.4654} \\ &= 130.3935 \end{aligned}$$

Wednesday
8/11/17

Shifting the Base Year:-

Shifting the base year means changing the given period (Year) of a series of Index Numbers and finding them into a new series based on some recent new Base Period. This shifting of the base year will be necessary due to the following reasons.

- When the base year is too old or too distant from the current period to make meaningful and valid comparisons. It should be noted that the base year should not be too far away from the given current year.
- If we want to compare the series of Index numbers with different base periods, to make quick and valid comparisons, both the series must be expressed with a common base period.
- The approximate method followed in shifting the Index numbers is as follows

$$\text{Recast Index Number of any year} = \frac{\text{Old Index No. of the year}}{\text{Index No. of New base year}} \times 100$$

1) Reconstruct the following Indices using 2000 as the base year

Year :	1996	1997	1998	1999	2000	2001	2002
Index NO :	110	130	150	175	180	200	220

Sol

Year	Index NO	Index Number (Base 2000)
1996	110	$\frac{110}{180} \times 100 = 61.1111$
1997	130	$\frac{130}{180} \times 100 = 72.2222$
1998	150	$\frac{150}{180} \times 100 = 83.3333$
1999	175	$\frac{175}{180} \times 100 = 97.2222$
2000	180	$\frac{180}{180} \times 100 = 100$
2001	200	$\frac{200}{180} \times 100 = 111.1111$
2002	220	$\frac{220}{180} \times 100 = 122.2222$

2) An Index is at 100 in 1981, it raises 4% in 1982, falls 6% in 1983, falls 4% in 1984 and raises 3% in 1985. Calculate the Index Numbers for the five years with 1983 as the base year.

Year	Index NO	Index number (Base 1983)
1981	100	$\frac{100}{97.76} \times 100 = 102.2913$
1982	104	$\frac{104}{97.76} \times 100 = 106.38$
1983	$\frac{94}{100} \times 104 = 97.76$	$\frac{97.76}{97.76} \times 100 = 100$
1984	$\frac{96}{100} \times 97.76 = 93.84$	$\frac{93.84}{97.76} \times 100 = 95.99$
1985	$\frac{103}{100} \times 93.84 = 96.65$	$\frac{96.65}{97.76} \times 100 = 98.86$

3) The following are the Price Index Numbers with 1985 as the Base Year

Year :	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
Index NO :	100	120	122	116	120	120	137	136	149	156

Shift the base to 1990 and Recast the Index Numbers.

Sol

<u>Year</u>	<u>Index No</u>	<u>Index Number (Base 1990)</u>
1985	100	$\frac{100}{120} \times 100 = 83.3333$
1986	120	$\frac{120}{120} \times 100 = 100$
1987	122	$\frac{122}{120} \times 100 = 101.6667$
1988	116	$\frac{116}{120} \times 100 = 96.67$
1989	120	$\frac{120}{120} \times 100 = 100$
1990	120	$\frac{120}{120} \times 100 = 100$
1991	137	$\frac{137}{120} \times 100 = 114.17$
1992	136	$\frac{136}{120} \times 100 = 113.33$
1993	149	$\frac{149}{120} \times 100 = 124.17$
1994	156	$\frac{156}{120} \times 100 = 130$
1995	137	$\frac{137}{120} \times 100 = 114.17$

4) An Index Number in 1997 is 100. It raises 15% in 1998, falls 4% in 1999, falls 2% in 2000 and raises 10% in 2001. Determine the Index Numbers for five years with 1999 as base year.

Sol

<u>Year</u>	<u>Index No</u>	<u>Index Number (1999)</u>
1997	100	$\frac{100}{110.4} \times 100 = 90.58$
1998	115	$\frac{115}{110.4} \times 100 = 104.17$
1999	$\frac{96}{100} \times 115 = 110.4$	$\frac{110.4}{110.4} \times 100 = 100$
2000	$\frac{98}{100} \times 110.4 = 108.1920$	$\frac{108.1920}{110.4} \times 100 = 98$
2001	$\frac{110}{100} \times 108.1920 = 119.0112$	$\frac{119.0112}{110.4} \times 100 = 107.80$

* Cost of Living Index Number:-

The Wholesale Price Index Numbers measure the changes in the General Level of Prices and they ^{explain} fail to reflect the effect of the increase or decrease of the Prices on the cost of living of different Groups of People in the Society.

* The cost of living Index Numbers are also termed as CONSUMER PRICE INDEX NUMBERS which are designed to measure the effects of changes in the Prices of goods and Services on the Purchasing power of People during any given current period with respect to the Base Period.

* Hence the cost of living Index Numbers are constructed separately for different classes of People, based on their geographical locations also like Urban areas, Rural Areas, Semi-Urban Areas and Metro Areas.

* It should be understood that the cost of living Index Number measure the changes in the cost of living (or) Purchasing Power of a Particular class of People, due to the rise or fall in the retail Prices.

* The cost of living Index numbers do not measure the changes in the cost of living as a consequence of the changes in the living standards. The cost of living Index Numbers are based on the retail prices and Price is considered as a factor which influences the Purchasing power of a class of the People. But the Prices of consumer goods is only one of the various factors on which the standard of living of people depends like family size, Income and occupation, place and region.

* The cost of living Index is also called as "consumer Price Index" or "Retail Price Index".

- The consumer Price Index is calculated taking into consideration the different commodities like

Food items — Wheat, Rice, Pulses, Meat, Milk Products, oils,

fruits and vegetables, Sugar etc.

⇒ clothing, fuel and lighting, House rent and Miscellaneous Expenses like Medical expenses, Education, ^(entertainment) Recreation, Transport, Personal care etc.

* The Relative Importance of Different items of consumption is different for different classes of people and even within the same class from Region to Region. Accordingly the cost of living index numbers are obtained as Weighted Indices, by taking into consideration the relative Importance of the commodities which is decided on the basis of the amount spent on various items.

* The Consumer Price Index number or cost of living Index number is calculated using aggregate expenditure method (Weighted Aggregate method)

$$\begin{aligned}\text{COST OF LIVING INDEX} &= \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100 \\ &= \frac{\text{Total Expenditure in current year}}{\text{Total Expenditure in Base year}} \times 100\end{aligned}$$

Where Total Expenditure in current year is obtained with base year quantities as weights

$$\text{Cost of Living Index Number} = \frac{\sum WI}{\sum W} = \frac{\sum WP}{\sum W}$$

$$I = \text{Price relative} = \frac{P_1}{P_0} \times 100$$

Uses:-

* The cost of living Index Numbers are used to determine the purchasing power of money and to calculate the real wages (income) from the nominal or money wages. The central and state Governments and other private organisations make use of the cost of living index numbers to calculate DA - Dearness Allowance (OR) Grant of bonus to etc. Employees to compensate them for increased cost of living due to price rise.

1) construct the cost of living Index Number from the following given table

Group	Index for 1998	Expenditure
1. Food	550	46%.
2. clothing	215	10%.
3. fuel & Lighting	220	7%.
4. House Rent	150	12%.
5. Miscellaneous	275	25%.

i) how is the overall Index Number altered if all the group index numbers are changed in the same ratio.

b) If all the weights and group indices are changed in the same ratio.

c) If all the group Index Numbers are increased by '10' and all the weights doubled.

Sol i) cost of Living Index number for 1998 = $\frac{\sum WI}{\sum W}$

Group =	Index for 1998	Expenditure	WI
Food	550	46%.	25300
clothing	215	10%.	2150
fuel & Lighting	220	7%.	1540
House Rent	150	12%.	1800
Miscellaneous	275	25%.	6875
		$\sum W = 100$	$\sum WI = 37665$

$$\Rightarrow \frac{\sum WI}{\sum W} = \frac{37665}{100} = 376.65$$

i) a) Suppose all the group Index Numbers are changed in the same ratio, say 'K', the new group Index number will be

$$\begin{aligned}\text{Cost of Living Index} &= \frac{\sum W(KI)}{\sum W} = Y \times K \frac{\sum WI}{\sum W} \\ &\Rightarrow Y \times K (376.65) \\ &\Rightarrow 376.65 K\end{aligned}$$

The overall cost of living Index Number also changes in the same ratio

b) Suppose all the group Indices are changed in the same ratio say 'a' and all weights are changed in the same ratio 'b', then the

$$\text{overall cost of living Index No} = \frac{\sum WI}{\sum W}$$

$$\Rightarrow \frac{\sum (bW)(aI)}{\sum bW}$$

$$\Rightarrow \frac{a \cancel{b} \sum WI}{\cancel{b} \sum W}$$

$$\Rightarrow \frac{a \sum WI}{\sum W}$$

$$\Rightarrow 376.65 a$$

i.e., the overall Index number changes in the same ratio as group Indices

c)

$$\text{cost of Living Index No} = \frac{\sum 2W(I+10)}{\sum 2W}$$

$$= \frac{2 \cancel{W} \sum (I+10)}{\sum 2W}$$

$$= \frac{\sum 2W(I+10)}{\sum 2W}$$

$$= \frac{\cancel{2} \sum W(I+10)}{\cancel{2} \sum W}$$

$$= \frac{\sum (WI + 10W)}{\sum W}$$

$$\Rightarrow \frac{\sum WI}{\sum W} + \frac{10 \sum W}{\sum W} \Rightarrow 376.65 + 10$$

$$= 386.65$$

- 2) In the construction of a certain cost of living Index number, the following group Index numbers were found. calculate the cost of living Index number by using
- The Weighted Arithmetic mean.
 - The Weighted Geometric mean.

Group	Index No	Weight
1. Food	350	5
2. Fuel & Lighting	200	1
3. clothing	240	1
4. House Rent	160	1
5. Miscellaneous	250	2

Group	Index I	Weight w	wI	log I	w log I
Food	350	5	1750	2.5441	12.7205
Fuel & Lighting	200	1	200	2.3010	2.3010
clothing	240	1	240	2.3802	2.3802
House Rent	160	1	160	2.2041	2.2041
Miscellaneous	250	2	500	2.3979	4.7958
		$\Sigma w = 10$	$\Sigma wI = 2850$	$\Sigma w \log I = 24.4016$	

i) The consumer Price Index No

using Arithmetic mean

$$\Rightarrow P_{01}(A.M) = \frac{\Sigma wI}{\Sigma w} = \frac{2850}{10} = 285$$

ii) using Geometric mean

$$\Rightarrow \log P_{01}(G.M) = \frac{\Sigma w \log I}{\Sigma w} = \frac{24.4016}{10}$$

$$\log P_{01}(G.M) = 2.44016$$

$$P_{01}(G.M) = \text{Antilog } 2.44016$$

$$= 275.5244$$

3) calculate the cost of living Index Number from the following data

Items	Price		weights
	Base year	current year	
Food	30	47	4
Fuel	8	12	1
clothing	14	18	3
House Rent	22	15	2
Miscellaneous	25	30	1

Sol

Items	Weights W	Prices		Price Relatives $I = P_1/P_0 \times 100$	WI
		Base year	current year		
Food	4	30	47	156.67	626.68
Fuel	1	8	12	150.00	150
clothing	3	14	18	128.57	385.71
House Rent	2	22	15	68.18	136.36
Miscellaneous	1	25	30	120.00	120.00
	11				1418.75

$$\text{Cost of Living Index NO} = \frac{\sum WI}{\sum W}$$

$$= \frac{1418.75}{11}$$

$$= 128.98$$

4) In the construction of a certain cost of Living Index Number, the following group Index numbers were found. Calculate the cost of living Index Number by using i) The weighted Arithmetic mean.
ii) The Weighted Geometric mean

Group	Index No	weights
Food	352	48
Fuel & Lighting	200	10
Clothing	230	8
House Rent	160	12
Other Miscellaneous	190	15

Group	Index Number I	Weights (w)	wI	$\log I$	$w \log I$
Food	352	48	16896	2.5465	122.2320
Fuel & Lighting	200	10	2000	2.3010	23.01
clothing	230	8	1840	2.3617	18.8936
House Rent	160	12	1920	2.2041	26.4492
Miscellaneous	190	15	2850	2.2788	34.1820
		93	25506		224.7668

Consumer Price Index Number

i) using Arithmetic Mean

$$\Rightarrow P_{01}(A.M) = \frac{\sum wI}{\sum w} = \frac{25506}{93}$$

$$= \frac{224.7668}{93} = 274.2581$$

ii) using Geometric Mean

$$\Rightarrow \log P_{01}(G.M) = \frac{\sum w \log I}{\sum w}$$

$$\log P_{01}(G.M) = \frac{224.7668}{93}$$

$$\log P_{01}(G.M) = 2.4168$$

$$P_{01}(G.M) = \text{Antilog } 2.4168$$

$$= 261.0959$$

Friday
10/11/17

5) In 1981 for working class people, wheat was selling at an average price of Rs. 16 per 10 kgs, cloth at Rs. 4 per meter, + house rent at Rs. 50 per house and Miscellaneous items at Rs. 20 per unit. By 1991 the cost of wheat increased by Rs. 8 per 10 kgs, + house rent by Rs. 30 per house and Miscellaneous items doubled the price. The weights of the groups in order were in the ratio of 1:2:4:1. The working class cost of living Index Number for the year 1991

With 1981 as the base was 175%. By how much the cloth price increased during the Period 1981-91.

Sol Let us suppose that the cloth price in 1991 was Rs. x per metre and let the weights for the four groups wheat, cloth, House rent and Miscellaneous be $k, 2k, 4k, k$ respectively, so that they are in the ratio 1:2:4:1

Group	unit	Price		Weight w	$I = \frac{P_1}{P_0} \times 100$	wI
		1981 P_0	1991 P_1			
wheat	10 kg	16	24	k	150	150k
cloth	1 metre	4	(increased $\times$ price)	$2k$	$25x$	$50xk$
House rent	1 house	50	80	$4k$	160	640k
Miscellaneous	1 unit	20	40	k	200	200k
				$\Sigma w = 8k$		$990k + 50xk$

$\Rightarrow$ Given that, Index number = 175

$$\Rightarrow \frac{\Sigma wI}{\Sigma w} = 175$$

$$\Rightarrow \frac{990k + 50xk}{8k} = 175$$

$$\Rightarrow \frac{k(990 + 50x)}{8k} = 175$$

$$\Rightarrow 990 + 50x = 175 \times 8$$

$$\Rightarrow 990 + 50x = 1400$$

$$\Rightarrow 50x = 410$$

$$\Rightarrow x = 8.2$$

The increased cost of per metre cloth = cost in 1991 - cost in 1981
 $= 8.2 - 4$

$$= \text{Rs. } 4.2$$

6) In calculating a certain cost of living index Number, the following weights were used. food - 15, clothing - 3, Rent - 4, fuel & Light - 2 and Miscellaneous - 1. calculate the Index for a

date when the average percentage increase in prices of items in the various groups over the base period were 32, 54, 47, 78 and 58 respectively. Suppose a business Executive was earning Rs. 2050 in the base period, what should be his salary in the current period if his standard of living is to remain the same.

Sol

Group	Average % increase in price	Group Index I	Weight (W)	WI
Food	32	132	15	1980
cloth	54	154	3	462
Rent	47	147	4	588
Fuel & Light	78	178	2	356
Miscellaneous	58	158	1	158
			25	3544

Cost of Living Index Number

$$= \frac{\sum WI}{\sum W} \Rightarrow \frac{3544}{25} = 141.76$$

It means that if a person was getting Rs. 100 in the base year, then in order to fully compensate the business executive for increase in the prices, his salary in the current period should be 141.76. Hence if the business executive was earning Rs. 2050 in the base period

$$\begin{aligned} 100 &\rightarrow 2050 \\ 141.76 &\rightarrow \frac{2050}{100} \times 141.76 \\ &= 2906.80 \end{aligned}$$

7) A Textile Worker in the city of Mumbai earns 3500 per month. The cost of Living Index for a Particular month is given as 136. using the following data, find out the amount he spent on clothing and house rent

Group	Expenditure	Group Index
Food	1400	180
clothing	?	150
House rent	?	100
Fuel & Light	560	110
Miscellaneous	630	80

Group	Expenditure (W)	Group Index (I)	WI
Food	1400	180	2,52,000
clothing	x	150	150x
House rent	y	100	100y
Fuel & Light	560	110	61,600
Miscellaneous	630	80	50,400
	$x+y+2590$	$\sum WI = 3,64,000 + 150x + 100y$	

We have

$$\text{cost of Living Index No} = \frac{\sum WI}{\sum W} = 136 (\text{given})$$

$$\Rightarrow \frac{364000 + 150x + 100y}{x+y+2590} = 136$$

$$\Rightarrow \frac{364000 + 150x + 100y}{3500} = 136$$

$$\Rightarrow 364000 + 150x + 100y = 4,76,000$$

$$\Rightarrow \boxed{150x + 100y = 1,12,000} \longrightarrow \textcircled{1}$$

$$x+y+2590 = 3500$$

$$\boxed{x+y = 910} \longrightarrow \textcircled{2}$$

Solving $\textcircled{1}$ & $\textcircled{2}$ multiplying Equation $\textcircled{2}$ with '100'

$$\begin{array}{r} 150x + 100y = 1,12,000 \\ 100x + 100y = 91,000 \\ \hline (-) \end{array}$$

$$50x = 21,000$$

$$x = 420$$

Substituting $x = 420$ in equation (2)

$$x + y = 910$$

$$\Rightarrow 420 + y = 910$$

$$\Rightarrow y = 910 - 420$$

$$\Rightarrow y = 490$$

$\therefore$ clothing (x) = 420 ; House Rent (y) = 490

8) from the following data relating to working class consumer Price Index of a city, calculate the Index number for the years 1998 and 1999.

Group	Food	clothing	fuel & Lighting	House Rent	Miscellaneous
Weight	48	18	7	13	14
Group Index 1998	110	120	110	100	110
Group Index 1999	130	125	120	100	135

The wages were increased by 8% in 1999. If this increase sufficient?

Sol

Group	Weight (w)	Group Index 1998 (I_1)	Group Index 1999 (I_2)	wI_1	wI_2
Food	48	110	130	5280	6240
clothing	18	120	125	2160	2250
fuel & Lighting	7	110	120	770	840
House Rent	13	100	100	1300	1300
Miscellaneous	14	110	135	1540	1890
	$\Sigma w = 100$			$\Sigma wI_1 = 11050$	$\Sigma wI_2 = 12520$

$$\text{Index Number for 1998} = \frac{\sum W I_1}{\sum W} = \frac{11050}{100} = 110.5$$

$$\text{Index Number for 1999} = \frac{\sum W I_2}{\sum W} = \frac{12520}{100} = 125.20$$

$$\text{Increase in the Index Number from 1998 to 1999} = 125.20 - 110.5 = 14.7$$

Percentage increase in the Price Index for 1999

$$= \frac{14.7}{110.5} \times 100$$

$$= 13.3\%$$

We observe that 8% increase in wages from 1998 to 1999 is less than the Percentage increase in the cost of living Index Number.

Hence Insufficient.

9) The following table gives the cost of living index number for 1998 with 1986 as base for different commodity groups. Food, clothing, fuel & light, Rent and Miscellaneous as 440, 500, 350, 400 and 250 respectively with their weights in the order in the ratio 15:1:2:3:4 obtain the overall cost of living index number. Suppose a Person was earning RS.4000/- in 1986. What should be his salary in 1998 to maintain the same standard of living as in 1986.

Commodity	Index Number I	Weight (W)	WI
Food	440	15x	6600x
Clothing	500	x	500x
Fuel & Light	350	2x	700x
Rent	400	3x	1200x
Miscellaneous	250	4x	1000x
		$\sum W = 25x$	$\sum WI = 10000x$

Cost of living Index Number for 1998 with respect to base year 1986

$$\frac{\sum WI}{\sum W} = \frac{10,000 \times}{25 \times} = 400$$

This means that if a person was getting Rs. 100/- in the base year (1986) in order that he can fully compensate for increase in the prices, his salary in the current period (1998) should be Rs. 400/-

Hence, if a person was earning Rs. 400/- in 1986 then his salary in 1998 would be

$$100 \rightarrow 4000$$

$$400 \rightarrow \frac{4000}{100} \times 400 = 16,000$$

To enable him to maintain the same standard of living with respect to the increase in prices, other factors remaining constant

Q. An enquiry into the budget of the middle class families of a certain city revealed on an average the percentage expenses on the different groups were food - 45, Rent - 15, clothing - 12, fuel & light - 8, Miscellaneous - 20. The group index number for the current year as compared with a fixed base period were 410, 150, 343, 248 and 285 respectively. Calculate the cost of living index number for the current year. Mr. X was getting Rs. 2400/- in the base period and Rs. 4300/- in the current year, state how much he need to receive as extra allowance to maintain his former standard of living.

Group	Weight (W)	Index Number (I)	WI
Food	45	410	18450
Rent	15	150	2250
Clothing	12	343	4116
Fuel & Light	8	248	1984
Miscellaneous	20	285	5700
	$\sum W = 100$		$\sum WI = 32500$

$$\text{Cost of Living Index number} = \frac{\sum WI}{\sum W}$$

$$= \frac{32800}{100}$$

$$= 325$$

The salary needed in the current year

$$100 \rightarrow 2400$$

$$325 \rightarrow \frac{2400}{100} \times 325 = 7800$$

But, Mr. X is receiving Rs. 4300 which is less than required salary 7800

So, to maintain same standard of living, the extra allowance needed $\Rightarrow 7800 - 4300$

$$= \text{Rs. } 3500/-$$

11) The data below shows the % increase in the price of a few selected food items and the weights attached to each of them. calculate the Index Number for the Food group.

Food items :	Rice	wheat	Dal	Ghee	oil	Spices	Milk	Fish
weight :	33	11	8	5	5	3	7	9
% increase in price :	180	202	115	212	175	517	260	426

Vegetables

9

332

Refreshment

10

279

Food items :	food	clothing	fuel & light	Rent & Rates	Miscellaneous
Index :	-	310	220	150	300
Weight :	60	5	8	9	18

Food items	Weight (W)	% increase	Current Index I	W I
Rice	33	180	(100+180) 280	9240
wheat	11	202	302	3322
Dal	8	115	215	1720
Ghee	5	212	312	1560
oil	5	175	275	1375
Spices	3	517	617	1851
milk	7	260	360	2520
Fish	9	426	526	4734
vegetables	9	332	432	3888
Re-freshment	10	279	379	3790
$\Sigma W = 100$				$\Sigma W I = 34000$

$$\text{Food Index} = \frac{\Sigma W I}{\Sigma W} = \frac{34000}{100} = 340$$

here, we find the Index number for food i.e. 340

Group	Index (I)	Weight (W)	W I
Food	340	60	20400
clothing	310	5	1550
Fuel & Light	220	8	1760
Rent & Rates	150	9	1350
Miscellaneous	300	18	5400
		$\Sigma W = 100$	$\Sigma W I = 30460$

$$\begin{aligned} \text{cost of living Index Number} &= \frac{\Sigma W I}{\Sigma W} \\ &= \frac{30460}{100} \\ &= 304.60 \end{aligned}$$

Monday
13/11/17
2017/2018

Question Paper Model

compute the fisher's Ideal Index Number from the data given below

Commodity	1990		2000	
	Price	quantity	price	quantity
A	8	50	20	60
B	2	15	6	10
C	1	20	2	25
D	2	10	5	8
E	1	40	5	30

Sol

Commodity	P_0	q_0	P_1	q_1	$P_0 q_0$	$P_1 q_1$	$P_0 q_1$	$P_1 q_0$
A	8	50	20	60	400	1200	480	1000
B	2	15	6	10	30	60	20	90
C	1	20	2	25	20	50	25	40
D	2	10	5	8	20	40	16	50
E	1	40	5	30	40	150	30	200
					510	1500	571	1380

$$\text{Fisher's Index Number} = \sqrt{L \times P}$$

$$L = \frac{\sum P_1 q_0}{\sum P_0 q_0} = \frac{1380}{510} = 2.7059$$

$$\text{Paasche Index} = \frac{\sum P_1 q_1}{\sum P_0 q_1} = \frac{1500}{571} = 2.6270$$

$$\text{Fisher's Index} = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}} \times 100$$

$$= \sqrt{2.7059 \times 2.6270 \times 100}$$

$$= \sqrt{7.1083 \times 100}$$

$$= 2.6661 \times 100$$

$$= 266.61$$

2016
2) Compute Fisher's Ideal Index Number from the data given below and check whether the Time reversal Test and factor Reversal Test are Satisfied

Commodity	Base year		Current year	
	price	Quantity	price	Quantity
A	2	7	6	6
B	3	6	2	3
C	4	5	8	5
D	5	4	2	4

Commodity	P ₀	q ₀	P ₁	q ₁	P ₀ q ₀	P ₁ q ₁	P ₁ q ₀	P ₀ q ₁
A	2	7	6	6	14	36	42	12
B	3	6	2	3	18	6	12	9
C	4	5	8	5	20	40	40	20
D	5	4	2	4	20	8	8	20
					72	90	102	61

$$\text{Fisher's Ideal Index Number} = \sqrt{\frac{\sum P_1 q_0 \times \sum P_0 q_1}{\sum P_0 q_0 \times \sum P_1 q_1}} \times 100$$

$$= \sqrt{\frac{102 \times 90}{72 \times 61}} \times 100$$

$$= \sqrt{\frac{9180}{4392}} \times 100$$

$$= \sqrt{2.0902} \times 100$$

$$= 1.4457 \times 100$$

$$= 144.57$$

Time Reversal Test:-

$$P_{01}^F \times P_{10}^F = 1$$

$$P_{01}^F = 144.57 = 1.4457$$

$$P_{10}^F = \sqrt{\frac{\sum P_0 q_1 \times \sum P_0 q_0}{\sum P_1 q_1 \times \sum P_1 q_0}} \times 100$$

$$= \sqrt{\frac{61 \times 72}{102 \times 90}} \times 100$$

$$= \sqrt{\frac{4392}{9180}} \times 100$$

$$= \sqrt{0.4784} \times 100$$

$$= 0.6917 \times 100$$

$$= 69.1687$$

Then

$$P_{01}^F \times P_{10}^F$$

$$= 1.4457 \times 0.6917$$

$$= 1$$

Then Time Reversal Test is Satisfied

Factor Reversal Test:-

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 q_1}{\sum P_0 q_0}$$

$$\text{Where } P_{01}^F = \sqrt{\frac{\sum P_1 q_0 \times \sum P_1 q_1}{\sum P_0 q_0 \times \sum P_0 q_1}} = 1.4457$$

$$Q_{01}^F = \sqrt{\frac{\sum P_0 q_1 \times \sum P_1 q_1}{\sum P_0 q_0 \times \sum P_1 q_0}} = \sqrt{\frac{61 \times 90}{102 \times 72}} = \sqrt{\frac{5490}{7344}} = \sqrt{0.7475}$$

$$= 0.8646$$

$$P_{01}^F \times Q_{01}^F = 1.4457 \times 0.8646$$

$$= 1.25$$

2015
3)

What

$$\frac{\sum P_1 Q_1}{\sum P_0 Q_0} = \frac{90}{72}$$

$$= 1.25$$

then

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

Factor Reversal Test is Satisfied.

2015
3)

What is Index Number. Explain Time Reversal and Factor Reversal Test.

2014
4)

With the help of the following data, Prove that the Fisher's Ideal Index Satisfies both Time Reversal and Factor Reversal Tests

commodity	Base year		current year	
	Price	Quantity	Price	Quantity
A	5	50	6	72
B	7	84	10	80
C	10	80	12	96
D	4	20	5	36

2014, 2014
5) Define Index Numbers and Explain its features. Compute Fisher's Ideal Index Number from the following data and show that it satisfies Time Reversal and factor Reversal Test

Item	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	12	10	15	12
B	15	7	20	5
C	24	5	20	9
D	5	16	5	14

2017
I)

Time Series Problems

Fit a Second degree Trend to the following data

Year : 1955 1960 1965 1970 1975 1980 1985

Production :- 6 8 9 10 12 11 8
('000 tons)

Year	X	Y	X ²	X ³	X ⁴	XY	X ² Y
1955	-15	6	225	-3375	50625	-90	1350
1960	-10	8	100	-1000	10000	-80	800
1965	-5	9	25	-125	625	-45	225
1970	0	10	0	0	0	0	0
1975	5	12	25	125	625	60	300
1980	10	11	100	1000	10000	110	1100
1985	15	8	225	3375	50625	120	1800
	0	64	700	0	$\Sigma X^4 = 122500$	$\Sigma XY = 75$	5575

The Normal Equations are

$$\Sigma y = na + b \Sigma x + c \Sigma x^2$$

$$\Sigma xy = a \Sigma x + b \Sigma x^2 + c \Sigma x^3$$

$$\Sigma x^2 y = a \Sigma x^2 + b \Sigma x^3 + c \Sigma x^4$$

$$\Sigma y = na + b \Sigma x + c \Sigma x^2$$

$$64 = 7a + b(0) + c(700)$$

$$7a = 64 - 700c \quad \neq 64$$

$$\boxed{7a + 700c = 64} \longrightarrow \textcircled{1}$$

$$\Sigma xy = a \Sigma x + b \Sigma x^2 + c \Sigma x^3$$

$$75 = a(0) + b(700) + c(0)$$

$$75 = 700b$$

$$b = \frac{75}{700} = 0.1071 \quad b = \frac{75}{700} = 0.1071$$

$$\boxed{b = 0.1071}$$

$$\boxed{b = 0.1071}$$

$$\Sigma x^2 y = a \Sigma x^2 + b \Sigma x^3 + c \Sigma x^4$$

$$5575 = a(700) + b(0) + c(122500)$$

$$\boxed{700a + 122500c = 5575} \longrightarrow \textcircled{2}$$

Solving $\textcircled{1}$ & $\textcircled{2}$ equations and multiplying equation $\textcircled{1}$ with 100

$$(7a + 700c = 64) \times 100$$

$$700a + 70000c = 6400$$

$$700a + 122500c = 5575$$

$$\begin{array}{r} (-) \quad (-) \quad (-) \\ \hline \end{array}$$

$$- 52500c = 825$$

$$c = -0.0157$$

$$\boxed{c = -0.0157}$$

Substituting $c = -0.0157$ in equation $\textcircled{1}$

$$7a + 700(-0.0157) = 64$$

$$7a - 71 = 64$$

$$7a = 64 + 11$$

$$7a = 75$$

$$a = 10.7143$$

2016

2) What is Time Series What is the need to analyse the Time Series.
Explain fitting of Second degree Equation.

2015

3/2011

The Sales of a Company in Thousands of Rupees for the years 1980 to 1984 are given below

Year : 1980 1981 1982 1983 1984

Sales : 32 47 65 92 132

Estimate the Sales for the year 1985 using $y = ab^x$

Sol

Year	Sales y	$\log y$	Time centered X	X^2	$X \log y$
1980	32	1.5051	-2	4	-6.0204
1981	47	1.6721	-1	1	-1.6721
1982	65	1.8129	0	0	0
1983	92	1.9638	1	1	1.9638
1984	132	2.1206	2	4	4.2411
	368	9.0744	0	10	1.5225

$$\log a = \frac{\sum \log y}{n} = \frac{9.0744}{50} = 1.8149$$

$$\log b = \frac{\sum X \log y}{\sum X^2} = \frac{1.5225}{10} = 0.1523$$

$$\Rightarrow y = ab^x$$

$$\Rightarrow \log y = \log a + x \log b$$

$$\Rightarrow y = \text{Antilog} (1.8149 + \text{Antilog}(0.15225))^x$$

$$y = (65.2980) (1.4199)^x$$

Sales on 1985 $x = 3$

$$y = (65.2980) (1.4199)^3 \Rightarrow (65.2980) (2.8625) = 186.4175$$

Year : 2001 2002 2003 2004 2005

Production: 7 9 10 7 5

fit a Parabolic Trend $y = a + bx + cx^2$

Year	<u>x</u>	<u>y</u>	<u>x²</u>	<u>x³</u>	<u>x⁴</u>	<u>x²y</u>	<u>xy</u>
2001	-2	7	4	-8	16	28	-14
2002	-1	9	1	-1	1	9	-9
2003	0	10	0	0	0	0	0
2004	1	7	1	1	1	7	7
2005	2	5	4	8	16	20	10
	<u>0</u>	<u>38</u>	<u>10</u>	<u>0</u>	<u>34</u>	<u>64</u>	<u>-6</u>

Normal Trends

$$\Sigma y = na + b \Sigma x + c \Sigma x^2$$

$$38 = 5a + b(0) + c(10)$$

$$\boxed{5a + 10c = 38} \rightarrow \textcircled{1}$$

$$\Sigma xy = a \Sigma x + b \Sigma x^2 + c \Sigma x^3$$

$$-6 = a(0) + b(10) + c(0)$$

$$10b = -6 \Rightarrow \boxed{b = -0.6}$$

$$\Sigma x^2 y = a \Sigma x^2 + b \Sigma x^3 + c \Sigma x^4$$

$$64 = a(10) + b(0) + c(34)$$

$$\boxed{10a + 34c = 64} \rightarrow \textcircled{2}$$

Solving equation $\textcircled{1}$ & $\textcircled{2}$ and multiplying Equation $\textcircled{1}$ with 2

$\textcircled{1} \times 2$

$$\begin{array}{r} 10a + 20c = 76 \\ 10a + 34c = 64 \\ \hline (-) \quad (-) \quad (-) \\ -14c = 12 \end{array}$$

$$c = \frac{-12}{14} \Rightarrow c = -0.8571$$

Substituting $c = -0.8571$ in equation ①

$$5a + 10(-0.8571) = 38$$

$$5a - 8.571 = 38$$

$$5a = 46.571$$

$$a = 9.3142$$

Substituting the values of a, b, c in the Non-linear Trend equation
We have

$$\hat{y} = a + bx + cx^2$$

$$\hat{y} = 9.3142 - 0.6x - 0.8571x^2$$

2001

$$\hat{y} = 9.3142 - 0.6(-2) - 0.8571(4)$$

$$\hat{y} = 9.3142 + 1.2 - 3.4284$$

$$\hat{y} = 7.0858$$

2002

$$\hat{y} = 9.3142 - 0.6(-1) - 0.8571(1)$$

$$\hat{y} = 9.3142 + 0.6 - 0.8571$$

$$\hat{y} = 9.0571$$

2003

$$\hat{y} = 9.3142 - 0.6(0) - 0.8571(0)$$

$$\hat{y} = 9.3142$$

2004

$$\hat{y} = 9.3142 - 0.6(1) - 0.8571(1)$$

$$\hat{y} = 7.8571$$

2005

$$\hat{y} = 9.3142 - 0.6(2) - 0.8571(4)$$

$$\hat{y} = 9.3142 - 1.2 - 3.4284$$

$$\hat{y} = 4.6858$$

²⁰¹³ 5) fit a straight line trend by the method of least square and estimate the Production in 2012

Year :	2006	2007	2008	2009	2010
Production:	18	21	23	27	16

Year	y	x	xy	x ²
2006	18	-2	-36	4
2007	21	-1	-21	1
2008	23	0	0	0
2009	27	1	27	1
2010	16	2	32	4
	105	0	2	10

$$\sum y = na + b \sum x$$

$$105 = 5a + b(0)$$

$$a = 21$$

$$\sum xy = a \sum x + b \sum x^2$$

$$2 = a(0) + b(10)$$

$$2 = 10b \Rightarrow b = \frac{2}{10} = 0.2$$

$$y = a + bx \Rightarrow y = 21 + 0.2x$$

2006

$$\hat{y} = 21 + 0.2(-2)$$

$$\hat{y} = 21 - 0.4 \Rightarrow \hat{y} = 20.6$$

2007

$$\hat{y} = 21 + 0.2(-1)$$

$$\hat{y} = 20.8$$

2008

$$\hat{y} = 21 + 0.2(0)$$

$$\hat{y} = 21$$

2009 $\hat{y} = 2$

$$\hat{y} = 21 + 0.2(1)$$

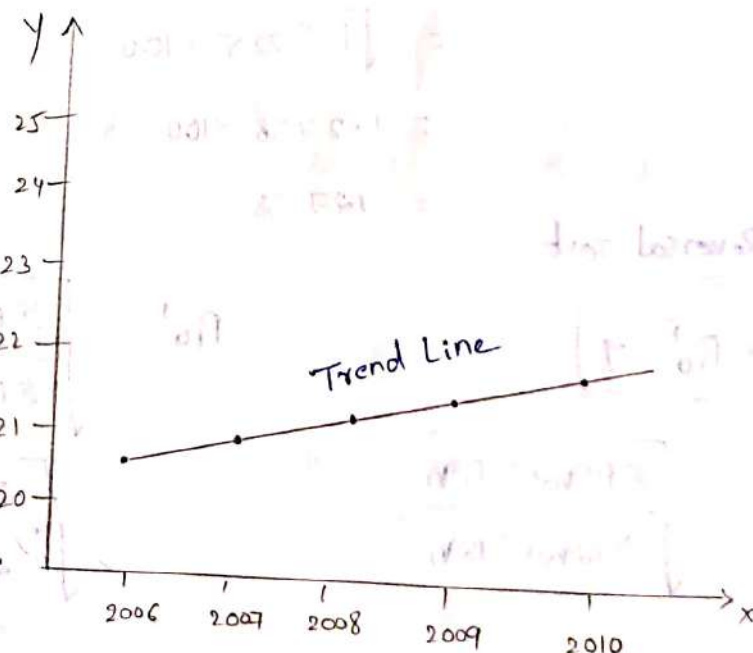
$$\hat{y} = 21.2$$

2010

$$\hat{y} = 21 + 0.2(2)$$

$$\hat{y} = 21 + 0.4$$

$$\hat{y} = 21.4$$



The production in 2012

$$x = 4$$

$$\Rightarrow y = 21 + 0.2x$$

$$y = 21 + 0.2(4)$$

$$y = 21 + 0.8$$

$$y = 21.8$$

Commodity	P_0	Q_0	P_1	Q_1	P_0Q_0	P_1Q_1	P_0Q_1	P_1Q_0
A	5	50	6	72	250	432	360	300
B	7	84	10	80	588	800	560	840
C	10	80	12	96	800	1152	960	960
D	4	20	5	30	80	150	120	100
					1718	2534	2000	2200

$$\begin{aligned}
 \text{Fisher's Index} &= \sqrt{\frac{\sum P_1Q_0 \times \sum P_0Q_1}{\sum P_0Q_0 \times \sum P_1Q_1}} \times 100 \\
 &= \sqrt{\frac{2200 \times 2534}{1718 \times 2000}} \times 100 \\
 &= \sqrt{\frac{5574800}{3436000}} \times 100 \\
 &= \sqrt{1.6225} \times 100 \\
 &= 1.2738 \times 100 \\
 &= 127.38
 \end{aligned}$$

Time Reversal Test

$$P_{01}^F \times P_{10}^F = 1$$

$$\begin{aligned}
 P_{01}^F &= \sqrt{\frac{\sum P_1Q_0 \times \sum P_0Q_1}{\sum P_0Q_0 \times \sum P_1Q_1}} \\
 &= 1.2738
 \end{aligned}$$

$$\begin{aligned}
 P_{10}^F &= \sqrt{\frac{\sum P_0Q_1 \times \sum P_1Q_0}{\sum P_1Q_1 \times \sum P_0Q_0}} \\
 &= \sqrt{\frac{2000 \times 1718}{2534 \times 2200}} \\
 &= \sqrt{\frac{3436000}{5574800}} \\
 &= \sqrt{0.6163} \\
 &= 0.7851
 \end{aligned}$$

$$\begin{aligned}
 P_{01}^F \times P_{10}^F &= (1.2738) (0.7851) \\
 &= 1
 \end{aligned}$$

Hence Time Reversal Test is Satisfied.

Factor Reversal Test:-

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$$

$$\text{Where } P_{01}^F = \sqrt{\frac{\sum P_1 Q_0 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_0 Q_1}}$$

$$= 1.2738$$

$$Q_{01}^F = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_1 Q_0}}$$

$$= \sqrt{\frac{2000 \times 2534}{2200 \times 1718}}$$

$$= \sqrt{\frac{5068000}{3779600}}$$

$$= \sqrt{1.3409}$$

$$= 1.1580$$

$$P_{01}^F \times Q_{01}^F = 1.2738 \times 1.1580$$

$$= 1.4750$$

Now

$$\frac{\sum P_1 Q_1}{\sum P_0 Q_0} = \frac{2534}{1718}$$

$$= 1.4750$$

$$\therefore P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} \quad \text{factor Reversal Test is satisfied}$$

Sol

Items	P_0	Q_0	P_1	Q_1	$P_0 Q_0$	$P_1 Q_1$	$P_0 Q_1$	$P_1 Q_0$
A	12	10	15	12	120	180	144	150
B	15	7	20	5	105	100	75	140
C	24	5	20	9	120	180	216	100
D	5	16	5	14	80	70	70	80
					425	530	505	470

$$\text{Fisher's Index} = \sqrt{L \times P} = \sqrt{\frac{\sum P_1 Q_0 \times \sum P_1 Q_1}{\sum P_0 Q_0 \times \sum P_0 Q_1} \times 100}$$

$$= \sqrt{\frac{470 \times 530}{425 \times 505}} \times 100$$

$$= \sqrt{\frac{249100}{214625}} \times 100$$

$$= \sqrt{1.1606} \times 100$$

$$= 1.0773 \times 100$$

$$= 107.73$$

Time Reversal Test:-

$$P_{01}^F \times P_{10}^F = 1$$

$$P_{01}^F = \sqrt{\frac{\sum P_1 Q_{01} \times \sum P_1 Q_{11}}{\sum P_0 Q_{01} \times \sum P_0 Q_{11}}}$$

$$= \sqrt{\frac{470 \times 530}{425 \times 505}}$$

$$= \sqrt{\frac{249100}{214625}}$$

$$= \sqrt{1.1606}$$

$$= 1.0773$$

$$P_{10}^F = \sqrt{\frac{\sum P_0 Q_{11} \times \sum P_0 Q_{01}}{\sum P_1 Q_{11} \times \sum P_1 Q_{01}}}$$

$$= \sqrt{\frac{505 \times 425}{470 \times 530}}$$

$$= \sqrt{\frac{214625}{249100}}$$

$$= \sqrt{0.8616}$$

$$= 0.9282$$

$$P_{01}^F \times P_{10}^F = 1.0773 \times 0.9282$$

$= 1$ Hence Time Reversal Test satisfied.

Factor Reversal Test

$$P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_{11}}{\sum P_0 Q_{11}}$$

Where

$$P_{01}^F = \sqrt{\frac{\sum P_1 Q_{01} \times \sum P_1 Q_{11}}{\sum P_0 Q_{01} \times \sum P_0 Q_{11}}}$$

$$= 1.0773$$

$$Q_{01}^F = \sqrt{\frac{\sum P_0 Q_1 \times \sum P_1 Q_0}{\sum P_0 Q_0 \times \sum P_1 Q_1}}$$

$$= \sqrt{\frac{505 \times 530}{425 \times 470}}$$

$$= \sqrt{\frac{267650}{199750}}$$

$$= \sqrt{1.3399}$$

$$= 1.1576$$

$$P_{01}^F \times Q_{01}^F = (1.0773)(1.1576)$$

$$= 1.2470$$

Now

$$\frac{\sum P_1 Q_1}{\sum P_0 Q_0} = \frac{530}{425}$$

$$= 1.2470$$

then $P_{01}^F \times Q_{01}^F = \frac{\sum P_1 Q_1}{\sum P_0 Q_0}$ Hence Factor Reversal Test satisfied.

Extry
14/11/17

PROBLEMS ON CORRELATION

2017
1) Calculate the Rank correlation co-efficient from the following data

X: 48 33 35 48 16 16 24 57 35
Y: 13 13 24 24 15 9 6 6 9

X	R ₁	Y	R ₂	d = R ₁ - R ₂	d ²
48	2.5	13	4.5	-2	4
33	6	13	4.5	1.5	2.25
35	4.5	24	1.5	3	9
48	2.5	24	1.5	1	1
16	8.5	15	3	5.5	30.25
16	8.5	9	6.5	2	4
24	7	6	8.5	-1.5	2.25
57	1	6	8.5	-7.5	56.25
35	4.5	9	6.5	-2	4

$$\sum d^2 = 113$$

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(113)}{9(9^2-1)}$$

$$= 1 - \frac{678}{720}$$

$$= 1 - 0.9416$$

$$= 0.0584$$

2) Find the correlation co-efficient to the following data

X 10 12 13 12 16

Y 40 38 43 45 37

Distinguish between correlation and Regression. Find the correlation to the given data.

X	Y	$x = x - \bar{x}$	$y = y - \bar{y}$	x^2	y^2	xy
10	40	-2.6	-0.6	6.76	0.36	1.56
12	38	-0.6	-2.6	0.36	6.76	1.56
13	43	0.4	2.4	0.16	5.76	0.96
12	45	-0.6	4.4	0.36	19.36	-2.64
16	37	3.4	-3.6	11.56	12.96	-12.24
63	203			19.2	45.20	-10.86

$$\bar{x} = \frac{\sum x}{n} = \frac{63}{5} = 12.6$$

$$\bar{y} = \frac{\sum y}{n} = \frac{203}{5} = 40.6$$

$$\text{Co-efficient of correlation} = \frac{\sum xy}{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}}$$

$$= \frac{-10.8}{\sqrt{19.2} \sqrt{45.2}}$$

$$= \frac{-10.8}{(4.3818)(6.7231)}$$

$$= \frac{-10.8}{29.4593}$$

$$= -0.3666$$

$$= \frac{18.96}{\sqrt{19.2} \sqrt{45.2}}$$

$$= \frac{18.96}{(4.3818)(6.7231)}$$

$$= \frac{18.96}{29.4593}$$

$$= 0.6436$$

3) Explain different types of correlation. calculate Karl Pearson co-efficient of correlation for the following data.

Roll No. of students	:	1	2	3	4	5
Marks in Accounts (x)	:	48	35	17	23	47
Marks in Statistics (y)	:	45	20	40	25	45

2012
4) i) Define correlation co-efficient. Explain Spearman's Rank correlation co-efficient

ii) calculate Rank correlation co-efficient to the following data

X:	61	72	73	63	73	84	80	66	76	74	72
Y:	40	52	59	53	61	58	56	42	58	50	50

2013
5) Explain different types of correlation. calculate Karl Pearson's co-efficient for the given data.

X:	10	12	14	16	20	22
Y:	9	11	5	13	4	2

2016
6) From the following data find the two Regression equations. Hence find the correlation co-efficient

X:	25	28	35	32	31	36	34	32
Y:	43	46	49	41	36	32	31	30

X	Y	$x - \bar{x}$ dx	$y - \bar{y}$ dy	dx^2	dy^2	$dx dy$
25	43	-6.625	4.5	43.8906	20.25	-29.83
28	46	-3.625	7.5	13.18	56.25	-27.22
35	49	3.4	10.5	11.56	110.25	35.7
32	41	0.37	2.5	0.1369	6.25	0.9250
31	36	-0.63	-2.5	0.3969	6.25	-1.5750
36	32	4.37	-6.5	19.0969	42.25	-28.40
34	31	2.37	-7.5	5.6169	56.25	-17.77
32	30	0.37	-8.5	0.1369	72.25	-3.14
253	308			94.0151	370	-68.16

$$\bar{x} = \frac{\sum x}{n} = \frac{253}{8} = 31.625$$

$$\bar{y} = \frac{\sum y}{n} = \frac{308}{8} = 38.5$$

Regression Equation

y on x

$$\Rightarrow y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$\Rightarrow y - \bar{y} = b_{yx} (x - \bar{x})$$

$$\Rightarrow y - 38.5 = (-0.7250)(x - 31.63)$$

$$\Rightarrow y - 38.5 = -0.7250x + 22.9314$$

$$\boxed{y = -0.7250x + 61.4314}$$

x on y

$$\Rightarrow (x - \bar{x}) = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$\Rightarrow (x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$(x - 31.63) = (-0.1842)(y - 38.5)$$

$$x - 31.63 = -0.1842y + 7.0923$$

$$\boxed{x = -0.1842y + 38.7223}$$

$$b_{yx} = \frac{\sum dxdy}{\sum x^2}$$
$$= \frac{-68.16}{94.0151}$$

$$= -0.7250$$

$$b_{xy} = \frac{\sum dxdy}{\sum dy^2}$$
$$= \frac{-68.16}{370}$$
$$= -0.1842$$

Correlation co-efficient $\Rightarrow r = \pm \sqrt{b_{yx} b_{xy}}$

$$= \pm \sqrt{(-0.7250)(-0.1842)}$$

$$= \pm \sqrt{0.1335}$$

$$= \pm 0.3654$$

$$\Rightarrow r = -0.3654$$

2017
7/ calculate the two Regression Lines from the following data

Sales : 87 93 104 117 63 120 47 69 107 53

Purchases : 74 78 72 100 73 94 43 64 84 51

2013
8) Determine the two Regression Lines from the following data

Annual Expenditure: 56 42 36 47 49 42 60 72 63 55
(Rs. '0000)

Annual Profit: 147 125 118 128 145 140 155 160 149 150

Also estimate the annual Profit when the Annual Expenditure

is Rs. 4,50,000

2012

9) Obtain the equations of two lines of Regression for the following data

A: 23 43 53 63 73 83

B: 5 6 7 8 9 10

2014

10) Distinguish between correlation and Regression, calculate the Regression Lines from the data given below

Price: 10 12 13 12 16 15

Amount: 40 38 43 45 37 43

Price X	Amount Y	$x - \bar{x}$ dx	$y - \bar{y}$ dy	dx^2	dy^2	$dx dy$
10	40	-3	-170.1667	9	28956.70	510.48
12	38	-1	-176.16	1	31032.34	176.16
13	43	0	-167.16	0	27942.46	0
12	45	-1	-165.16	1	27227.82	165.16
16	37	3	-173.16	9	29984.38	-519.48
15	43	2	-167.16	4	27942.46	-334.32
78	246			24	173136.16	-2

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

$$= \frac{78}{6}$$

$$= 13$$

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

$$= \frac{246}{6}$$

$$= 41.01667$$

Regression Equations

Y on X

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

$$= y - \bar{y} = b_{yx} (x - \bar{x})$$

$$b_{yx} = \frac{\sum dx dy}{\sum x^2}$$

$$= \frac{-2}{24}$$

$$= -0.0833$$

$$y - 210.16 = (-0.0833)(x - 13)$$

$$y - 210.16 = -0.0833x + 1.0829$$

$$\boxed{y = -0.0833x + 211.2429}$$

x on y

$$\Rightarrow (x - \bar{x}) = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$\Rightarrow (x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$x - 13 = (-0.0095)(y - 210.16)$$

$$x - 13 = -0.0095y + 2$$

$$\boxed{x = -0.0095y + 15}$$

$$b_{xy} = \frac{\sum dxdy}{\sum y^2}$$

$$= \frac{-2}{210.16}$$

$$= -0.0095$$

The Regression Lines are

$$y = -0.0833x + 211.2429$$

$$x = -0.0095y + 15$$

The co-efficient of correlation is

$$r = \pm \sqrt{b_{xy} / b_{yx}}$$

$$r = \pm \sqrt{(-0.0095)(-0.0833)}$$

$$r = \pm \sqrt{0.0008}$$

$$r = -0.0281$$

3sol

X	Y	$x - \bar{x}$ $\bar{x}$	$y - \bar{y}$ $\bar{y}$	x^2	y^2	xy
48	45	14	10	196	100	140
35	20	1	-15	1	225	-15
17	40	-17	5	289	25	-85
23	25	-11	-10	121	100	-110
47	45	13	10	169	100	130
170	175			776	550	280

$$\bar{X} = \frac{\sum x}{n} = \frac{170}{5} = 34$$

$$\bar{y} = \frac{\sum y}{n} = \frac{175}{5} = 35$$

Correlation co-efficient of Karl Pearson = $r = \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$

$$r = \frac{280}{\sqrt{776} \sqrt{550}}$$

$$r = \frac{280}{(27.8568)(23.4521)}$$

$$r = \frac{280}{653.2999}$$

$$r = 0.4286$$

4 Sol

X	R ₁	Y	R ₂	R ₁ - R ₂ d	d ²
61	11	40	11	0	0
72	7.5	52	7	0.5	0.25
73	5.5	59	2	3.5	12.25
63	10	53	6	4	16
73	5.5	61	1	4.5	20.25
84	1	58	3.5	-2.5	6.25
80	2	56	5	-3	9
66	9	42	10	-1	1
76	3	58	3.5	-0.5	0.25
74	4	50	8.5	-4.5	20.25
72	7.5	50	8.5	-1	1
					86.50

Rank correlat co-efficient

$$= 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(86.50)}{11(11^2-1)}$$

$$= 1 - \frac{519}{1320}$$

$$= 1 - 0.3932$$

$$= 0.6068$$

Sol

X	y	$x - \bar{x}$	$y - \bar{y}$	x^2	y^2	xy
10	9	-5.67	1.67	32.1489	2.7889	-9.4689
12	11	-3.67	3.67	13.4689	13.4689	-13.4689
14	5	-1.67	-2.33	2.7889	5.4289	3.8911
16	13	0.33	5.67	0.1089	32.1489	1.8711
20	4	4.33	-3.33	18.7489	11.0889	-14.4189
22	2	6.33	-5.33	40.0689	28.4089	-33.7389
94	44			107.3334	93.3334	-65.3334

$$\bar{x} = \frac{\sum x}{n} = \frac{94}{6}$$

$$= 15.6667$$

$$; \bar{y} = \frac{\sum y}{n} = \frac{44}{6}$$

$$= 7.3333$$

Karl Pearson co-efficient $\Rightarrow r = \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$

$$r = \frac{-65.3334}{\sqrt{(107.3334)} \sqrt{(93.3334)}}$$

$$r = \frac{-65.3334}{(10.3602) (9.6609)}$$

$$r = \frac{-65.3334}{100.0891}$$

$$r = -0.6528$$

7 Sol

Sales X	Purchases Y	$x - \bar{x}$ $\frac{dx}{dx}$	$y - \bar{y}$ $\frac{dy}{dy}$	dx^2	dy^2	$dx dy$
87	74	1	0.7	1	0.49	0.7
93	78	7	4.7	49	22.09	32.9
104	72	18	-1.3	324	1.69	-23.4
117	100	31	26.7	961	712.89	827.70
63	73	-23	-0.3	529	0.09	6.9
120	94	34	20.7	1156	428.49	703.80
47	43	-39	-30.3	1521	918.09	1181.7
69	64	-17	-9.3	289	86.49	158.10
107	84	21	10.7	441	114.49	224.7
53	51	-33	-22.3	1089	497.29	735.9
860	733			6360	2782.10	3849

$$\bar{x} = \frac{\sum x}{n} = \frac{860}{10} = 86 ; \quad \bar{y} = \frac{\sum y}{n} = \frac{733}{10} = 73.3$$

Regression Equations

y on x

$$\Rightarrow (y - \bar{y}) = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$\Rightarrow (y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$\Rightarrow y - 73.3 = 0.6052 (x - 86)$$

$$\Rightarrow y - 73.3 = 0.6052x - 52.0462$$

$$\Rightarrow y = 0.6052x - 52.0462 + 73.3$$

$$\Rightarrow \boxed{y = 0.6052x + 21.2538}$$

$$b_{yx} = \frac{\sum dx dy}{\sum dx^2} = \frac{3849}{6360} = 0.6052$$

x on y

$$\Rightarrow (x - \bar{x}) = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$(x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$b_{xy} = \frac{\sum dx dy}{\sum dy^2} = \frac{3849}{2782.10} = 1.383$$

$$x - 86 = (1.3835)(y - 73.3)$$

$$x - 86 = 1.3835y - 101.4096$$

$$x = 1.3835y - 15.4096$$

Correlation co-efficient $r = \pm \sqrt{b_{xy} b_{yx}}$

$$= \pm \sqrt{(1.3835)(0.6052)}$$

$$= \pm \sqrt{0.8373}$$

$$= 0.9150$$

8sol

X	Y	$x - \bar{x}$ dx	$y - \bar{y}$ dy	dx^2	dy^2	$dx dy$
56	147	3.8	5.3	14.44	28.09	20.14
42	125	-10.2	-16.7	104.04	278.89	170.34
36	118	-16.2	-23.7	262.44	561.69	383.94
47	128	-5.2	-13.7	27.04	187.69	71.24
49	145	-3.2	3.3	10.24	10.89	-10.56
42	140	-10.2	-1.7	104.04	2.89	17.34
60	155	7.8	13.3	60.84	176.89	103.74
72	160	19.8	18.3	392.04	334.89	362.34
63	149	10.8	7.3	116.64	53.29	78.84
55	150	2.8	8.3	7.84	68.89	23.24
522	1417			1099.6	1704.3	1220.6

$$\begin{aligned}\bar{x} &= \frac{\sum x}{n} \\ &= \frac{522}{10} \\ &= 52.2\end{aligned}$$

$$\begin{aligned}\bar{y} &= \frac{\sum y}{n} \\ &= \frac{1417}{10} \\ &= 141.7\end{aligned}$$

Regression equations

y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 141.7 = 1.11(x - 52.2)$$

$$y - 141.7 = 1.11x - 57.942$$

$$y = 1.11x + 83.7559$$

$$\begin{aligned}b_{yx} &= \frac{\sum dx dy}{\sum dx^2} \\ &= \frac{1220.6}{1099.6} \\ &= 1.11\end{aligned}$$

x on y

$$\Rightarrow x - \bar{x} = b_{xy} (y - \bar{y})$$

$$\Rightarrow x - 52.2 = (0.7162)(y - 73.1417)$$

$$\Rightarrow x - 52.2 = 0.7162y - 101.4839$$

$$\Rightarrow \boxed{x = 0.7162y - 49.2839}$$

Annual profit when $x = 45$

y on x

$$y = 1.11(45) + 83.7559$$

$$y = 49.95 + 83.7559$$

$$\boxed{y = 133.7059}$$

9.50

x	y	$x - \bar{x}$ dx	$y - \bar{y}$ dy	dx^2	dy^2	$dx dy$
23	5	-33.33	-2.5	1110.88	6.25	83.3250
43	6	-13.33	-1.5	177.68	2.25	19.9950
53	7	-3.33	-0.5	11.08	0.25	1.6650
63	8	6.67	0.5	44.48	0.25	3.3350
73	9	16.67	1.5	277.88	2.25	25.0050
83	10	26.67	2.5	711.28	6.25	66.6750
338	45			2333.28	17.50	200

$$\bar{x} = \frac{\sum x}{n} \Rightarrow \frac{338}{6} = 56.33 \quad ; \quad \bar{y} = \frac{\sum y}{n} = \frac{45}{6}$$

$$= 7.5$$

Regression Lines

y on x

$$\Rightarrow y - \bar{y} = b_{yx} (x - \bar{x})$$

$$\Rightarrow y - 7.5 = 0.0857 (x - 56.33)$$

$$b_{yx} = \frac{\sum dx dy}{\sum x^2}$$

$$= \frac{200}{2333.28}$$

$$= 0.0857$$

$$y - 7.5 = 0.0857x - 4.8284$$

$$y = 0.0857x + 2.6716$$

x on y

$$(x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$x - 56.33 = (11.4286)(y - 7.5)$$

$$x - 56.33 = 11.4286y - 85.7143$$

$$x = 11.4286y - 29.3843$$

$$b_{xy} = \frac{\sum dxdy}{\sum y^2}$$

$$= \frac{200}{17.5}$$

$$= 11.4286$$

10 Sol

X	Y	$x - \bar{x}$ dx	$y - \bar{y}$ dy	dx^2	dy^2	$dxdy$
10	40	-3	-1	9	1	3
12	38	-1	-3	1	9	3
13	43	0	2	0	4	0
12	45	-1	4	1	16	-4
16	37	3	-4	9	16	-12
15	43	2	2	4	4	4
78	246			24	50	-6

$$\bar{x} = \frac{\sum x}{n} = \frac{78}{6} = 13$$

$$\bar{y} = \frac{\sum y}{n} = \frac{246}{6} = 41$$

$$b_{yx} = \frac{\sum dxdy}{\sum x^2} = \frac{-6}{24} = -0.25$$

$$b_{xy} = \frac{\sum dxdy}{\sum y^2} = \frac{-6}{50} = -0.12$$

y on x

$$y - \bar{y} = b_{yx} (x - \bar{x})$$

$$y - 41 = -0.25(x - 13)$$

$$y - 41 = -0.25x + 3.25$$

$$y = -0.25x + 44.25$$

x on y

$$x - \bar{x} = b_{xy} (y - \bar{y})$$

$$x - 13 = (-0.12)(y - 41)$$

$$x - 13 = -0.12y + 4.92$$

$$x - 13 = -0.12y + 17.92$$

$$x = -0.12y + 17.92$$

11) Distinguish b/w correlation and Regression. What are the Regression co-efficient state their properties. Explain multiple Regression and its significance with Applications.

2015
12) calculate correlation co-efficient from the following data

X : 0 1 2 3 4 5 6

Y : 2 1 3 2 4 3 5

2015
13) The following figures relate to Advertising Expenditure and Sales

Adv (Lakhs).
Expenditure : 60 62 65 70 73 75

Sales : 12 14 11 13 16 19
(crores)

Estimate the sales for Advertising Expenditure of Rs. 90,00,000

X	Y	$x - \bar{x}$	dx	dx^2	dy^2	$dx dy$
60	12	-7.5	-2.17	56.25	4.7089	16.27
62	14	-5.5	-0.17	30.25	0.03	0.93
65	11	-2.5	-3.17	6.25	10.04	7.92
70	13	2.5	-1.17	6.25	1.37	-2.92
73	16	5.5	1.83	30.25	3.35	10.06
75	19	7.5	4.83	56.25	23.33	36.22
405	85			185.5	42.83	68.48

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

$$= \frac{405}{6}$$

$$= 67.5$$

$$\bar{Y} = \frac{\sum y}{n}$$

$$= \frac{85}{6}$$

$$= 14.17$$

y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$b_{yx} = \frac{\sum dx dy}{\sum dx^2}$$

$$= \frac{68.48}{185.5}$$

$$= 0.3692$$

$$y - 14.17 = (0.3692)(x - 67.5)$$

$$y - 14.17 = 0.3692x - 24.9186$$

$$y = 0.3692x - 10.7486$$

For $x = 95$

$$y = 0.3692(95) - 10.7486 \Rightarrow y = 35.0740 - 10.7486$$

$$y = 24.33 \text{ Crores}$$

UNIT-I

2016

- i) Explain the construction of a frequency distribution table. What is the difference between a frequency distribution and cumulative frequency distribution?
- ii) Calculate the Mean Deviation for the following data

Wages	50-59	60-69	70-79	80-89	90-99	100-109	110-119
No. of Workers	15	40	50	60	45	90	15

Wages	frequency	mid point (m)	mf	deviation from $\bar{x}$ (d)	f d
50-59	15	54.5	817.5	-33.01	495.15
60-69	40	64.5	2580	-23.01	920.4
70-79	50	74.5	3725	-13.01	650.5
80-89	60	84.5	5070	-3.01	180.6
90-99	45	94.5	4252.5	6.99	314.55
100-109	90	104.5	9405	16.99	1529.1
110-119	15	114.5	1717.5	26.99	404.85
	315		27567.5		4495.15

$$\begin{aligned}\bar{x} &= \frac{\sum mf}{\sum f} \\ &= \frac{27567.5}{315} \\ &= 87.51\end{aligned}$$

$$\begin{aligned}\text{Mean Deviation} &= \frac{\sum f|d|}{\sum f} \\ &= \frac{4495.15}{315} \\ &= 14.2703\end{aligned}$$

2017

- 2) Define statistics and discuss its limitations. (Explain the methods of constructing i) Histogram ii) Ogive curve)

2015

- 3) Calculate the Mean Deviation from the following data

Age (years)	10-19	20-29	30-39	40-49	50-59	60-69
frequency	3	61	223	137	53	19

4) Compare Mean, Median and Mode

5) What is primary data. Explain the sources of primary data

By calculate Mean Deviation from the following data.

Pay Scales (in '000) : Less than (<10) <20 <30 <40 <50 <60 <70

frequency : 5 11 19 31 46 51 54

Scales	f	m	mf	$m - \bar{x}$	$f d $
0 - 10	5	5	25	-29.63	148.15
10 - 20	6	15	90	-19.63	117.78
20 - 30	8	25	200	-9.63	77.04
30 - 40	12	35	420	0.37	4.44
40 - 50	15	45	675	10.37	155.55
50 - 60	5	55	275	20.37	101.85
60 - 70	3	65	195	30.37	91.11
	54		1870		695.92

$$\bar{x} = \frac{\sum f}{N}$$

$$= \frac{54}{1}$$

$$\bar{x} = \frac{\sum mf}{\sum f}$$

$$= \frac{1870}{54}$$

$$= 34.63$$

$$\text{Mean Deviation} = \frac{\sum f|d|}{\sum f}$$

$$= \frac{695.92}{54}$$

$$= 12.8874$$

2012
6) The Lives of two models of Refrigerators in a recent year are as follows

Life : 0-2 2-4 4-6 6-8 8-10 10-12
(No. of years)

No. of Refrigerators:

Model A - 5 16 13 7 5 4

Model B - 2 7 12 19 9 1

Which model has greater uniformity

sol calculating co-efficient of variation for Model A Refrigerator

$$CV = \frac{\sigma}{\mu} \times 100$$

$$\text{Mean} = \frac{\sum fm}{N}$$

Model A

Life	Number	m	fm
0-2	5	1	5
2-4	16	3	48
4-6	13	5	65
6-8	7	7	49
8-10	5	9	45
10-12	4	11	44
	50		256

$$\begin{aligned} (\mu) \text{Mean} &= \frac{256}{50} \\ &= 5.12 \end{aligned}$$

Standard Deviation

d/width of the class

Life	No. of (f) Refrigerators	mid Point	m - μ (mean) d	$d' = d / \frac{\text{width of class}}{2}$	d'^2	fd'	fd'^2
0-2	5	1	-4.12	-2.06	4.24		
2-4	16	3	-2.12	-1.06	1.12		
4-6	13	5	-0.12	-0.6	0.36		
6-8	7	7	1.88	0.94	0.88		
8-10	5	9	3.88	1.94	3.76		
10-12	4	11	5.88	2.94	8.64		

Life	No. of Refrigerators (f)	mid point	(m-d)	d' = d/mid pt	fd'	-fd'	fd' ²
0-2	5	1	-5	-2.5	-12.5	-12.5	31.25
2-4	16	3	-3	-1.5	-24	-24	36
4-6	13	5	-1	-0.5	-6.5	-6.5	3.25
6-8	7	7	1	0.5	3.5	3.5	1.75
8-10	5	9	3	1.5	7.5	7.5	11.25
10-12	4	11	5	2.5	10	10	25
	<u>50</u>	<u>36</u>				<u>-22</u>	<u>108.5</u>

$$\text{Mean of Mid values} = \frac{36}{6} = 6$$

Standard Deviation

$$= \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2 \times C}$$

$$= \sqrt{\frac{108.5}{50} - \left(\frac{-22}{50}\right)^2 \times 2}$$

$$= \sqrt{2.17 - 0.1936 \times 2}$$

$$= \sqrt{1.9764 \times 2}$$

$$= 1.4058 \times 2$$

$$\sigma = 2.81$$

$$C_{VA} = \frac{\sigma}{\mu} \times 100$$

$$= \frac{2.81}{8.12} \times 100$$

$$= 54.88 \%$$

Model - B

Number Life	Life Number	m	fm
2	0-2	1	2
7	2-4	3	21
12	4-6	5	60
19	6-8	7	133
9	8-10	9	81
1	10-12	11	11
<u>SD</u>			<u>308</u>

$$\text{Mean} = \frac{\sum fm}{\sum f}$$

$$= \frac{308}{50}$$

$$\mu = 6.16$$

Standard Deviation

Life	No. of (f)	Midpoint (m)	d	d'	d ²	fd'	fd' ²
0-2	2	1	-5	-2.5	6.25	-5	12.5
2-4	7	3	-3	-1.5	2.25	-10.5	15.75
4-6	12	5	-1	-0.5	0.25	-6	3
6-8	19	7	1	0.5	0.25	9.5	4.75
8-10	9	9	3	1.5	2.25	13.5	20.25
10-12	1	11	5	2.5	6.25	2.5	6.25
<u>SD</u>		<u>36</u>				<u>4</u>	<u>62.5</u>

$$\text{Mid Values} = \frac{36}{6} = 6$$

$$d = m - \frac{\sum m}{N}$$

$$d' = d/2$$

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2 \times c}$$

$$= \sqrt{\frac{62.5}{50} - \left(\frac{4}{50}\right)^2 \times 2}$$

$$= \sqrt{1.25 - 0.064} \times 2$$

$$= \sqrt{1.186} \times 2$$

$$= 2(1.089)$$

$$\sigma = 2.178$$

$$\text{Co-efficient of Variation} = \frac{\sigma}{\mu} \times 100$$

$$= \frac{2.2303}{6.16} \times 100$$

$$= 0.3621 \times 100$$

$$= 36.21\%$$

Model B Refrigerator is having more uniformly than Model A.

Sol
3)

Age	frequency (f)	m	mf	d	f d
10-19	3	14.5	43.5	-24.6976	74.0928
20-29	61	24.5	1494.5	-14.6976	896.5536
30-39	223	34.5	7693.5	-4.6976	1047.5648
40-49	137	44.5	6096.5	5.3024	726.4288
50-59	53	54.5	2888.5	15.3024	811.0272
60-69	19	64.5	1225.5	25.3024	480.7456
	<u>496</u>		<u>19442</u>		<u>4036.4128</u>

$$\begin{aligned} \bar{x} &= \frac{\sum fm}{\sum f} \\ &= \frac{19442}{496} = 39.1976 \end{aligned}$$

$$\begin{aligned} \text{Mean Deviation} &= \frac{\sum f|d|}{\sum f} \\ &= \frac{4036.4128}{496} \\ &= 8.1379 \end{aligned}$$

7) Distinguish between Primary and Secondary data. What are the sources of collecting them.

calculate the co-efficient of variation to the following data

Series A — 4 8 4 15 10 11 9

Series B — 12 8 3 15 6 4 10

Series - A

x	$x - \mu$	$(x - \mu)^2$
4	-4.71	22.19
8	-0.71	0.504
4	-4.71	22.19
15	6.29	39.56
10	1.29	1.66
11	2.29	5.24
9	0.29	0.084
<u>61</u>		<u>91.43</u>

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

$$\mu = \frac{\sum x}{n}$$

$$= \frac{61}{7} = 8.71$$

$$\Rightarrow \sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

$$= \sqrt{\frac{91.43}{7}}$$

$$= \sqrt{13.06}$$

$$= 3.61$$

$$\Rightarrow CV = \frac{\sigma}{\mu} \times 100$$

$$= \frac{3.61}{8.71} \times 100$$

$$= 41.45\%$$

Series-B

x	$x - \mu$	$(x - \mu)^2$
12	3.71	13.7641
8	-0.29	0.0841
3	-5.29	27.9841
15	6.71	45.0241
6	-2.29	5.2441
4	-4.29	18.4041
<u>10</u>	<u>1.71</u>	<u>2.92</u>
<u>58</u>		<u>113.4246</u>

$$\mu = \frac{\sum x}{n}$$

$$\mu = \frac{58}{7}$$

$$= 8.29$$

$$\Rightarrow \sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

$$= \sqrt{\frac{113.4246}{7}}$$

$$= \sqrt{16.2035} = 4.0254$$

$$\Rightarrow CV = \frac{\sigma}{\mu} \times 100$$

$$= \frac{4.0254}{8.29} \times 100$$

$$= 48.56$$

Q8) What is Tabulation. What are its uses. What important points should be kept in mind while presenting a good statistical Table. What are Ogive curves.

(Q1)

find the co-efficient of variation from the following data

Marks in statistics : Above 0 10 20 30 40 50 60 70 80

No. of students = 150 140 100 80 80 70 30 14 0

Marks in Stats	No. of student f	Midpoint (m)	$d_i = d/100$	d	d_i^2	$f d_i$	$f d_i^2$
0 - 10	10 (150-140)	5	-3.5	-35	12.25	-35	122.5
10 - 20	40 (140-100)	15	-2.5	-25	6.25	-100	250
20 - 30	20 (100-80)	25	-1.5	-15	2.25	-30	45
30 - 40	0	35	-0.5	-5	0.25	0	0
40 - 50	10	45	0.5	5	0.25	5	2.5
50 - 60	40	55	1.5	25	2.25	60	90
60 - 70	16	65	2.5	35	6.25	40	100
70 - 80	14	75	3.5	12.25	49	171.5	
	150					-11	781.5

$$\text{Midpoint (d)} = \frac{5 + 15 + 25 + 35 + 45 + 55 + 65 + 75}{8} = \frac{320}{8} = 40$$

Marks	No. of Students (f)	mid point (m)	mf
0-10	10	5	50
10-20	40	15	600
20-30	20	25	500
30-40	0	35	35
40-50	10	45	450
50-60	40	55	2200
60-70	16	65	1040
70-80	14	75	1050
	150		5925

$$\text{Mean} = \frac{\sum mf}{\sum f} = \frac{5925}{150} = 39.5$$

Standard deviation

$$\sigma_c = \sqrt{\frac{\sum f d_i^2}{n} - \left(\frac{\sum f d_i}{n}\right)^2 \times C}$$

$$= \sqrt{\frac{781.5}{150} - \left(\frac{-11}{150}\right)^2 \times 10}$$

$$\sigma_c = 22.8136$$

$$C_v = \frac{\sigma}{\mu} \times 100 = \frac{22.8136}{39.5} \times 100$$

$$= 57.756$$

9/ calculate the co-efficient of Variation

Value : More than 800 > 700 > 600 > 500 > 400 > 300 > 200 > 100

Frequency : 14 44 96 175 381 527 615 660

Value	frequency	m	mf
100 - 200	45 (660-615)	150	6750
200 - 300	88 (615-527)	250	22000
300 - 400	146 (527-381)	350	51100
400 - 500	206 (381-175)	450	92700
500 - 600	79	550	43450
600 - 700	82	650	33800
700 - 800	30	750	22500
800 - 900	14	850	11900
	660		284200

$$\text{Mean } (\mu) = \frac{\sum mf}{\sum f}$$

$$= \frac{284200}{660}$$

$$= 430.6061$$

Standard Deviation :-

Value	frequency	m	d	$d' = d/100$	d'^2	fd'	fd'^2
100 - 200	45	150	-350	-3.5	12.25	-157.5	551.25
200 - 300	88	250	-250	-2.5	6.25	-220	550
300 - 400	146	350	-150	-1.5	2.25	-219	328.5
400 - 500	206	450	-50	-0.5	0.25	-103	51.5
500 - 600	79	550	50	0.5	0.25	39.5	19.75
600 - 700	52	650	150	1.5	2.25	78	117
700 - 800	30	750	250	2.5	6.25	75	187.5
800 - 900	14	850	350	3.5	12.25	49	171.5
	660	4000				-458	1482.197

$$d = m - \frac{\sum m}{n}$$

$$\text{midvalues } m = \frac{\sum m}{n}$$

$$= \frac{4000}{8}$$

$$= 500$$

$$\text{Standard Deviation } \sigma = \sqrt{\frac{\sum fd'^2}{n} - \left(\frac{\sum fd'}{n}\right)^2} \times C$$

$$= \sqrt{\left(\frac{1977}{660}\right) - \left(\frac{-458}{660}\right)^2} \times 100$$

$$= \sqrt{2.9955 - (-0.6939)^2} \times 100$$

$$= \sqrt{2.9955 - 0.4816} \times 100$$

$$= \sqrt{2.5139} \times 100$$

$$= 1.5855 \times 100$$

$$= 158.55$$

$$= \sqrt{2.2455 - (-0.6939)^2}$$

$$= \sqrt{2.2455 - 0.4816}$$

$$= \sqrt{1.7639}$$

$$= 1.3281$$

$$\text{Co-efficient of Variation } = \frac{\sigma}{\mu} \times 100$$

$$= \frac{1.3281}{430.6061} \times 100$$

$$= 0.0031 \times 100$$

$$CV = 0.31$$

$$= \frac{158.55}{430.6061} \times 100$$

$$= 0.3682 \times 100$$

$$CV = 36.8205$$

10) calculate the median and standard deviation

Marks :	> 0	> 10	> 20	> 30	> 40	> 50	> 60	> 70	> 80
No. of Students :	100	98	95	90	80	50	35	23	18

Sol

Marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of students:	2	3	5	10	30	15	12	5	18

Marks	No. of students (f)	Cummulative frequency
0-10	2	2
10-20	3	5
20-30	5	10
30-40	10	20
40-50	30	50
50-60	15	65
60-70	12	77
70-80	5	82
80-90	18	100
	<u>100</u>	

$$\text{Mean} = \frac{n+1}{2} = \frac{100+1}{2}$$

$$= \frac{101}{2} = 50.5$$

$$\text{Median} = \frac{n}{2} = \frac{100}{2} = 50$$

$$\text{Median} = l_1 + \frac{l_2 - l_1}{f} (m - c)$$

$$= 40 + \frac{50 - 40}{30} (50.5 - 20)$$

$$= 40 + \frac{10}{30} (30.5)$$

$$= 40 + 10.1666$$

$$= 50.1666$$

$$\text{Median} = 50.1666$$

Marks	No. of Students	Mid point	d	$d_i = d/10$	d_i^2	fd_i	fd_i^2
0-10	2	5	-40	-4	16	-8	32
10-20	3	15	-30	-3	9	-9	27
20-30	5	25	-20	-2	4	-10	20
30-40	10	35	-10	-1	1	-10	10
40-50	30	45	0	0	0	0	0
50-60	15	55	10	1	1	15	15
60-70	12	65	20	2	4	24	48
70-80	5	75	30	3	9	15	45
80-90	18	85	40	4	16	72	288
		405				89	485

$$\text{Mid point (d)} = \frac{405}{9} = 45$$

Standard Deviation :-

$$\sigma_c = \sqrt{\frac{\sum fd_i^2}{n} - \left(\frac{\sum fd_i}{n}\right)^2} \times C$$

$$= \sqrt{\frac{485}{100} - \left(\frac{89}{100}\right)^2} \times 10$$

$$= \sqrt{4.85 - 0.89} \times 10$$

$$= \sqrt{3.96} \times 10$$

$$= 1.99 \times 10$$

$$= 19.9$$

$$\approx 20$$

$$\sigma_c = 19.9$$

2013
11)

calculate the Variance of the following distribution

Profit (RS. lakhs) : Below 5 Below 10 Below 15 Below 20 Below 25

No. of companies : 2 7 14 27 48

Below 30
64

Below 35
75

Sol

Profit: - 0-5 5-10 10-15 15-20 20-25 25-30 30-35

f : 2 5 7 13 21 16 11

Profit	freq	mid point	d	d/c d/5	d ²	-fd	-fd ²
0-5	2	2.5	-15	-3	9	-6	18
5-10	5	7.5	-10	-2	4	-10	20
10-15	7	12.5	-5	-1	1	-7	7
15-20	13	17.5	0	0	0	0	0
20-25	21	22.5	5	1	1	21	21
25-30	16	27.5	10	2	4	32	64
30-35	11	32.5	15	3	9	33	99
	75	122.5				63	229

$$\text{Midpoint } d = \frac{122.5}{7} = 17.5$$

$$\sigma_c = \sqrt{\frac{\sum f d_i^2}{n} - \left(\frac{\sum f d_i}{n}\right)^2 \times c}$$

$$= \sqrt{\frac{229}{75} - \left(\frac{63}{75}\right)^2 \times 5}$$

$$= \sqrt{3.0533 - (0.84)^2 \times 5}$$

$$= \sqrt{3.0533 - 0.7056 \times 5}$$

$$= \sqrt{2.3477 \times 5}$$

$$= 1.5322 \times 5$$

$$= 7.661$$

$$\text{Variance } (\sigma^2) = (7.661)^2 = 58.6925$$

2014
12) calculate standard deviation from the following data

Bursting pressure (lb) : 5.0-9.9 10.0-14.9 15-19.9 20-24.9 25-29.9 30-?

No. of bags : 2 9 29 54 11 5

2015

13) calculate Standard Deviation - from the following data

Wages (Rs) : less than 10 < 20 < 30 < 40 < 50 < 60 < 70 < 80

No. of Workers : 12 30 65 107 157 202 222 230

2015

14) calculate Mean Deviation from the following data

Age (years) : 10-19 20-29 30-39 40-49 50-59 60-69

frequency : 3 61 223 137 53 19

Age	frequency	Midpoint	m.f	deviation from $\bar{x}$	f d
10-19	3	14.5	43.5	-24.6975	74.0925
20-29	61	24.5	1494.5	-14.6975	896.5475
30-39	223	34.5	7693.5	-4.6975	1047.5425
40-49	137	44.5	6096.5	5.3025	726.4425
50-59	53	54.5	2888.5	15.3025	811.0325
60-69	19	64.5	1225.5	25.3025	480.7475
	496	237	19442		4036.4

$$\text{Mean } \bar{x} = \frac{\sum mf}{\sum f} = \frac{19442}{496} = 39.1975$$

$$\text{Mean Deviation} = \frac{\sum f|d|}{N} = \frac{4036.4}{496} = 8.1379$$

2016

15) calculate Mean Deviation from the following data

Wages (weekly) : 50-59 60-69 70-79 80-89 90-99 100-109 110-119

frequency : 15 40 50 60 45 90 15

Wages	Frequency	Mid	m.f	d-from $\bar{x}$	f d
50-59	15	54.5	817.5	-33.0158	495.237
60-69	40	64.5	2580	-23.0158	920.632
70-79	50	74.5	3725	-13.0158	650.79
80-89	60	84.5	5070	-3.0158	180.948
90-99	45	94.5	4252.5	6.9842	314.289
100-109	90	104.5	9405	16.9842	1528.578
110-119	15	114.5	1717.5	26.9842	404.763
	<u>315</u>		<u>27567.5</u>		<u>4495.237</u>

$$\text{Mean} = \frac{\sum m.f}{N}$$

$$= \frac{27567.5}{315} = 87.5158$$

$$\text{Mean Deviation} = \frac{\sum f|d|}{N} = \frac{4495.237}{315} = 14.2705$$

Monday
20/11/19

UNIT - IV PROBABILITY

Probability is the chance that a Particular event will occur. What is the chance of getting a head when a coin is tossed? What is the chance of picking up a Particular card from a deck of playing cards? A company has launched a new Advertisement, what is the chance that it will be Successful?

Managers often base their decisions on an Analysis of uncertainty such as

1) What are the chances that the Sales will increase if the Prices decrease.

2) What is the chance that a new Investment will be profitable.

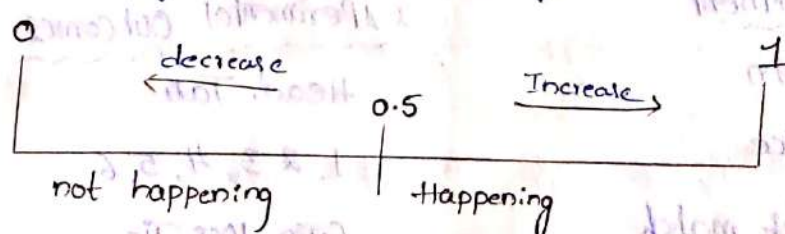
3) What is the chance that all the students of Ist Sem MBA will Pass in Distinction.

DEFINITION:-

Probability is defined as likelihood (or) chance that a Particular event will occur.

Probability is a numerical measure of likelihood that an event will occur. Thus Probabilities can be used as a measure of the degree of uncertainty.

Probability values are assigned on a Scale 0 to 1.



BASIC CONCEPTS:-

An Experiment is an Activity that results in

"one and only one outcome" out of a set of disjoint outcomes where

An outcome cannot be predicted with certainty.

Consider an experiment of tossing a die, there are six (6) outcomes and we cannot predict which outcome will come up when the die is thrown. In such cases the probability is $\frac{1}{6}$ ^{Experiment} _{Sample Space}.

→ Another example is tossing coin here there are two possible outcomes head and tail and excluding the possibilities of the coin standing on its edge. Suppose the two coins are tossed.

There are 4 possible outcomes (HH), (HT), (TT), (TH). These 4 outcomes can be denoted as 4 events represented as E_1, E_2, E_3 & E_4 .

Sample Space

The sample space is a set of possible outcomes of an experiment.

for Ex: If you have to toss a fair coin, then the sample space is $S = \{\text{Head, Tail}\}$. When a dice is thrown the sample space is $\{1, 2, 3, 4, 5, 6\}$. If in an experiment of drawing a card from a deck of playing cards the sample space is $S = 52$.

Therefore the sample space 's' for an experiment is set of all experimental outcomes.

Experiment

Toss a coin

Roll a dice

Play a cricket match

Launch of a product

If price of a

product increases

Experimental Outcomes

Head, Tail

1, 2, 3, 4, 5, 6

Win, Loss, Tie

Success and failure

Increase, decrease,
Remaining Constant.

FORMULAE :-

1. Each elementary event (or) a combination of elementary event must have associated with it a Probability greater than (or) equal to zero but less than (or) equal to one i.e., $0 \leq P(A) \leq 1$
2. The Probability of entire sample space i.e. if S represents sample space then $P(S) = 1$
3. The Probability that one (or) other (or) both of two mutually exclusive events will occur, is equal to the sum of individual Probabilities of these events.

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

4. The Probability of an event that does not occur is equal to one minus the Probability of the event that occurs.

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

Where $P(\bar{A})$ = Probability of Non Occurrence of an event.

$P(A)$ = Probability of Occurrence of an event.

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

$$\Rightarrow P(\bar{A}) + P(A) = P(S)$$

$$\frac{P}{1} = P(S)$$

Addition Law :-

The Addition Law for the Union of two events A & B is the event containing all sample points belonging to A or B or both i.e. it is denoted by $A \cup B$.

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

Multiplication Law :-

As the Addition Law of Probability is used to compute the Probability of a Union of two events the multiplication law is used to compute the Probability of Intersection of two events.

$$(i.e.) P(A \cap B) = P(B) \cdot P(A|B)$$

$$P(A \cap B) = P(A) \cdot P(B|A)$$

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

1. Assuming that a card is randomly selected from a deck of 52 playing cards, find the probability that
- The card drawn is king.
 - A card drawn is diamond.

Sol: (i): The card drawn is king:
 Number of samples $P(S) = 52$

$$\text{Probability of getting king } P(A) = 4 \Rightarrow \frac{P(A)}{P(S)} = \frac{4}{52} = \frac{1}{13}$$

(ii). A card drawn is a diamond:
 $P(S) = 52$

$$\text{No. of diamond in a deck } P(A) = 13 \Rightarrow \frac{P(A)}{P(S)} = \frac{13}{52} = \frac{1}{4}$$

2. Determine the probability of atleast one head appearing in two tosses, for tossing two coins.

Sol: No. of sample space $P(S) = 4$ (H,H), (T,H), (H,T), (T,T)

$$\text{Probability of getting atleast one head } P(A) = 3 \Rightarrow \frac{P(A)}{P(S)} = \frac{3}{4}$$

3. The word Professors containing some letters, what is the probability that it is a vowel.

Sol: sample space $P(S) = 9$

$$\text{Probability of getting Vowels } P(A) = 3 \Rightarrow \frac{P(A)}{P(S)} = \frac{3}{9} = \frac{1}{3}$$

4. A box contains 8 white balls, 6 Red balls and 4 green balls. find the probability that when they are drawn in the order (white, green & Red).

(i). When the ball is replaced after the draw.

(ii). When the ball is not replaced after the draw.

Sol: (i). When the ball is replaced after the draw

$$\text{Sample space } P(S) = 8 + 6 + 4 = 18$$

$$P(W, G, R) = \frac{8}{18} \times \frac{4}{18} + \frac{6}{18} \times \frac{8}{18} = \frac{8}{243} \approx 0.0329 \approx 0.033$$

When the ball is not replaced sample space $n(s) = 18$

$$P(W, G, R) = \frac{1}{18} \times \frac{1}{17} \times \frac{2}{16} = \frac{2}{51} = 0.039$$

5. An urn containing 5 white, 3 black & 2 red balls. If 3 balls are drawn at random, find the probability that i) all the 3 balls are white, (ii) 3 black, (iii) 2 one white and 1 is red, (iv) 2 are black & 1 is red (v). 1 ball of each colour is drawn.

Sol: Sample space $n(s) = 10$, 3 balls are drawn from urn

(i). All 3 balls are white: $P(W) = \frac{5}{18} \times \frac{4}{17} \times \frac{3}{16} = \frac{1}{12}$

(ii). All 3 balls are black: $P(B) = \frac{3}{18} \times \frac{2}{17} \times \frac{1}{16} = \frac{1}{120}$

(iii) 2 are white & 1 is Red: $P(W, R) = \frac{5}{18} \times \frac{4}{17} \times \frac{2}{16} = \frac{1}{18}$

(iv). 2 black & 1 Red: $P(B, R) = \frac{3}{18} \times \frac{2}{17} \times \frac{2}{16} = \frac{1}{20}$

(v). 1 ball of each colour: $P(W, B, R) = \frac{5}{18} \times \frac{3}{17} \times \frac{2}{16} = \frac{1}{8}$

- 6) A bag contains 4 white balls and 2 black balls. Another bag contains 3 white balls and 5 Black balls. If one ball is drawn from each bag, find the probability that i) Both are white ii) Both are Black iii) One ball is white and another ball is black

Sol Total balls in bag ① $\Rightarrow$ white - 4

Black - 2

6

Total balls in bag ② $\Rightarrow$ white - 3

Black - 5

8

i) Both are white

$$\frac{4}{6} \times \frac{3}{8} = \frac{1}{4} = 0.25$$

ii) The Probability that both are Black

$$\frac{2}{6} \times \frac{5}{8} = 0.208$$

ii) The Probability that one ball is white and another ball is black means it may be noted that the first ball drawn from first bag may be a white ball and the second ball drawn from the second bag may be a black ball and the opposite may also happen. Hence the probability is calculated as follows.

$$\Rightarrow P(W_1)P(B_2) + P(B_1)P(W_2)$$

$$\Rightarrow \left(\frac{4}{6} \times \frac{5}{8}\right) + \left(\frac{2}{6} \times \frac{3}{8}\right)$$

$$\Rightarrow 0.4167 + 0.1250$$

$$\Rightarrow 0.5417$$

7) Three contractors A, B and C are bidding for a contract. It is believed that A has exactly half the chance that B has. B in turn has $\frac{4}{5}$ th as likely as C has to win the contract. What is the probability for each contractor to win the contract.

Sol Let 'x' be the probability of C to get a contract

$$P(C) = x$$

$$\Rightarrow P(B) = \frac{4}{5}x$$

$$\Rightarrow P(A) = \frac{1}{2}\left(\frac{4}{5}x\right)$$

Now, we have $P(S) = 1$

$$\Rightarrow P(A) + P(B) + P(C) = 1$$

$$\Rightarrow \frac{2x}{5} + \frac{4x}{5} + x = 1$$

$$\Rightarrow \frac{2x + 4x + 5x}{5} = 1$$

$$\Rightarrow 2x + 4x + 5x = 5$$

$$\Rightarrow 11x = 5$$

$$\Rightarrow x = \frac{5}{11}$$

$$\Rightarrow x = 0.45$$

$$\boxed{P(C) = 0.45}$$

$$\Rightarrow P(B) = \frac{4}{5} \times$$

$$= \frac{4}{5} \times 0.45$$

$$P(B) = 0.36$$

$$\Rightarrow P(A) = \frac{2}{5} \times$$

$$= \frac{2}{5} \times 0.45$$

$$P(A) = 0.18$$

8) A Candidate is selected in an Interview for Management Trainee post in three companies. for the first company there are 12 candidates for the second company there are 15 candidates and for the third company there are 10 candidates. What are the chances of the candidate getting a job in atleast one of the companies.

Sol The Probabilities of the candidate getting a job in each of the companies

$$1^{\text{st}} \text{ company} = 1/12$$

$$2^{\text{nd}} \text{ company} = 1/15$$

$$3^{\text{rd}} \text{ company} = 1/10$$

The Probabilities that the candidate is not getting job in any of the companies

$$1^{\text{st}} \text{ company} = 1 - \frac{1}{12} = 11/12$$

$$2^{\text{nd}} \text{ company} = 1 - 1/15 = 14/15$$

$$3^{\text{rd}} \text{ company} = 1 - 1/10 = 9/10$$

The Probability that the candidate is not getting job in any of the companies

$$\frac{11}{12} \times \frac{14}{15} \times \frac{9}{10} = 0.77$$

The probability that the candidate will be selected in atleast one of the companies

1- Probability of the candidate not getting selected in any of the company

$$= 1 - 0.77$$

$$= 0.23$$

9) A Problem in statistics is given to 5 students of 1st sem M.B.A A, B, C, D, E. Their chances of solving are $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$, $\frac{1}{6}$ respectively. What is the probability that the problem will be solved.

Sol Probability of Any student not solving

$$P(A) = \frac{1}{2} \Rightarrow P(\bar{A}) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(B) = \frac{1}{3} \Rightarrow P(\bar{B}) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$P(C) = \frac{1}{4} \Rightarrow P(\bar{C}) = 1 - \frac{1}{4} = \frac{3}{4}$$

$$P(D) = \frac{1}{5} \Rightarrow P(\bar{D}) = 1 - \frac{1}{5} = \frac{4}{5}$$

$$P(E) = \frac{1}{6} \Rightarrow P(\bar{E}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Probability that no student can solve

$$\frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5} \times \frac{5}{6} \Rightarrow \frac{1}{6} = 0.1667$$

Probability that problem may be solved

$$\Rightarrow P(x) + P(\bar{x}) = P(s)$$

$$\Rightarrow P(\bar{x}) = 0.1667$$

$$\Rightarrow P(x) = 1 - 0.1667$$

$$\Rightarrow P(x) = 0.833$$

Thursday
23/11/17

10) A bag contains 3 red, 6 white and 7 blue balls. What is the probability that two balls drawn are white and blue

Sol

$$\frac{{}^2P_2}{{}^{16}P_2} \times \frac{7}{185} = \frac{7}{40} = 0.175$$

The probability of getting two balls white and blue is 0.175

11) A bag contains 7 white and 9 black balls. Two balls are drawn in succession at random what is the probability that one of them is white and the other is black

$$\left(\frac{7}{16} \times \frac{9}{15}\right) + \left(\frac{9}{16} \times \frac{7}{15}\right) = \frac{21}{80} + \frac{21}{80}$$

$$\left(\frac{7}{16} \times \frac{9}{15} = \frac{7}{16} \times \frac{3}{5} = \frac{21}{80} = 0.2625\right) = 0.2625 + 0.2625 = 0.5250$$

12) A bag contains 5 white and 3 black balls. Two balls are drawn at random one after other without replacement. find the probability that the balls drawn are black

Sol

$$\frac{3}{8} \times \frac{2}{7} = \frac{3}{28} = 0.1071$$

13) A bag contains 8 balls, 5 being red and 3 white. If a man selects two balls at random from the bag, what is the probability that he will get one ball of each colour.

Sol

$$\left(\frac{5}{8} \times \frac{3}{7}\right) + \left(\frac{3}{8} \times \frac{5}{7}\right) = \frac{15}{56} + \frac{15}{56}$$

$$= 0.2679 + 0.2679$$

$$= 0.5358$$

14) A bag contains 4 white, 5 red and 6 Green balls. Three balls are drawn at random. what is the probability that a white, a red, and a Green ball are drawn

Sol

$$\frac{4}{15} \times \frac{5}{14} \times \frac{6}{13} = \frac{12}{273}$$

$$= 0.0440$$

15) A bag contains 8 black, three red and 9 white balls. if

3 balls are drawn at random, find the probability that

i) All are black = $\frac{8}{20} \times \frac{7}{19} \times \frac{6}{18} = \frac{14}{285} = 0.0491$

ii) 2 are black and one is white = $\frac{8}{20} \times \frac{7}{19} \times \frac{9}{18} = \frac{14}{190} = 0.0737$

iii) one is of each colour = $\frac{8}{20} \times \frac{3}{19} \times \frac{9}{18} = \frac{6}{190} = 0.0316$

16) An urn contains 10 white and 6 black balls. Find the probability that the blind person in one draw shall obtain a white ball and in the second draw (without replacing the first one) a black ball.

Sol

$$\left(\frac{10}{16} \times \frac{6}{15} \right) = \frac{4}{4} = \frac{1}{4} = 0.25$$

17) An urn contains 10 white and 3 red balls. A ball is drawn and then replaced. Then another ball is drawn. What is the probability that a white and a red ball are drawn in that order.

Sol

$$\frac{10}{13} \times \frac{3}{13} = \frac{30}{169} = 0.1775$$

28/11/17

18) 3 Sales men A, B and C have been given a target of selling 10,000 units of a particular product. The Probability of them achieving their targets being respectively 0.25, 0.30 and 0.50. If these three sales men try to sell the product, find the probability of success of only one sales man. And the failure of the other two.

Sol

$$P(A) = 0.25 ; P(B) = 0.30 ; P(C) = 0.50$$

If 'A' is successful B & C are unsuccessful

$$P = P(A) [(1 - P(B)) (1 - P(C))]$$

$$= 0.25 [(1 - 0.3) (1 - 0.50)]$$

$$= 0.25 [(0.70) (0.50)]$$

$$= 0.0875$$

If B is successful A & C are unsuccessful

$$P = P(B) [(1 - P(A)) (1 - P(C))]$$

$$= 0.30 [(1 - 0.25) (1 - 0.50)]$$

$$= 0.30 [(0.75) (0.50)]$$

$$= 0.30 (0.3750)$$

$$= 0.1125$$

If 'C' is successful, B & A are unsuccessful

$$P = P(C) [(1 - P(A)) (1 - P(B))]$$

$$= 0.50 [(1 - 0.25) (1 - 0.30)]$$

$$= 0.50 [(0.75) (0.70)]$$

$$= 0.525 \times 0.50$$

$$= 0.2625.$$

∴ The required Probability on of them succeeded and the other two fails is

$$\Rightarrow P = 0.0875 + 0.1125 + 0.2625$$

$$P = 0.4625.$$

19) There are three brands x, y and z of a product in the market. A consumer chooses only one of them at a time. The Probabilities that he buys brand x, brand y and brand z are 0.20; 0.16; 0.45 respectively

i) What is the Probability that he doesn't buy any of the brands.

ii) Given that the consumer buys some brand, what is the Probability that he buys brand x (or) brand y (or) brand z.

Sol $P(X) = 0.20$; $P(Y) = 0.16$; $P(Z) = 0.45$

i) Probability of the consumer not purchasing any of the product

$$= 1 - \text{Probability of purchasing any of brands}$$

$$= 1 - [0.20 + 0.16 + 0.45]$$

$$= 1 - 0.81$$

$$= 0.19$$

ii) Probability of x $\Rightarrow \frac{P(X)}{P(X) + P(Y) + P(Z)} = \frac{0.20}{0.81} = 0.2469.$

probability of y $\Rightarrow \frac{P(Y)}{P(X) + P(Y) + P(Z)} = \frac{0.16}{0.81} = 0.1975$

probability of z $\Rightarrow \frac{P(Z)}{P(X) + P(Y) + P(Z)} = \frac{0.45}{0.81} = 0.55.$

20> An MBA graduate applies for a job in two companies X and Y. The probability of he being selected in a company X is 0.7 and being rejected in company Y is 0.5. The probability of at least one his applications being rejected is 0.6. What is the probability that he will be selected in one of the companies.

Sol The Probability of his selection in X company.

$$P(X) = 0.7$$

The Probability of his rejection in X company

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

$$= 1 - 0.70$$

$$P(\bar{X}) = 0.30$$

The Probability of his rejection in Y company

$$P(\bar{Y}) = 1 - 0.5 \Rightarrow 0.5 = 1 - P(Y)$$

$$P(Y) = 1 - 0.5$$

$$P(Y) = 0.5$$

The Probability of selection of any of the companies

$$= P(X) + P(Y) - (1 - P(\bar{X} \text{ or } \bar{Y}))$$

$$= 0.7 + 0.5 - (1 - 0.6)$$

$$= 1.2 - 0.4$$

$$= 0.8$$

9035587999

087161280

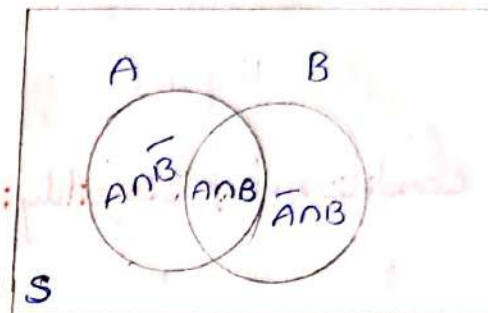
Theories

① State and Prove the Addition Theorem of Probability.

Statement:- The probability of occurrence of at least one of the two events A & B is given by

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

Proof:- Consider a random experiment with sample space 'S' with 'N' sample points (exhaustive event)



By the definition $P(A \cup B) = \frac{n(A \cup B)}{n(S)}$

$$\begin{aligned} P(A \cup B) &= \frac{n(A \cup B)}{n(S)} \\ &= \frac{n(A \cup B)}{N} \end{aligned}$$

Where, $n(A \cup B)$ is the no. of occurrences favourable to the events $A \cup B$

$$P(A \cup B) = \frac{n(A) - n(A \cap B) + n(A \cap B) + n(B) - n(A \cap B)}{N}$$

$$= \frac{n(A) + n(B) - n(A \cap B)}{N}$$

$$= \frac{n(A)}{N} + \frac{n(B)}{N} - \frac{n(A \cap B)}{N}$$

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

NOTE:- If the events A & B are mutually exclusive (Disjoint)

we have $A \cap B = \phi$,

$$\text{We have } n(A \cap B) = \frac{n(A \cap B)}{N} = \frac{n(\phi)}{N}$$

$$= 0$$

i.e. for mutually exclusive events A & B the addition theorem is

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

Suppose if there are three events A, B and C which are mutually exclusive

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

Mul Multiplication Theorem: [conditional Probability]

(OR)
Theorem of Compound Probability

Conditional Probability:-

Two events are said to be independent if the happening of an event is not effected by the happening of the other event. In other words happening of an event does not depend upon the happening of the other.

EX:- If we draw a card from a pack of 52 cards and replace it before we draw a second card, the second draw is independent of the first one. If don't replace the first card before the second draw, the second draw depends upon the first draw.

If 'B' is any arbitrary event in a sample space and $P(B) > 0$, The probability of A given B is defined by

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

where $P(A/B)$ is called "The conditional Probability" of 'A' given 'B'.

$$\Rightarrow P(A/B) = \frac{\text{no. of element in } (A \cap B)}{\text{no. of element in 'B'}}$$

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

$$P(B) \cdot P(A/B) = P(A \cap B)$$

Similarly $P(B \cap A) = P(B) \cdot P(A/B)$

Statement:- The Probability of Simultaneous happening of two events A & B is given by

$$P(A \cap B) = P(A) \cdot P(B/A) ; \text{ where } P(A) \neq 0$$

(or)

$$P(B \cap A) = P(B) \cdot P(A/B) ; \text{ where } P(B) \neq 0.$$

Where $P(B/A)$ is the conditional Probability of ~~two events~~ happening of B under the condition that ' A ' has happened and $P(A/B)$ is the conditional probability of happening of A under the condition that B has happened.

Proof:- The Probability of Simultaneous happening of two events A & B is the product of the Probabilities namely the probability of the first event times the conditional Probability of the second event, given that, the first event has already occurred. We may take any one other events A and B as the first event

Let the A & B be the events associated with the Sample Space ' S ' of a random Experiment with (Total) Exhaustive number of outcomes (Sample points) N i.e

$$n(S) = N$$

$$\Rightarrow P(A \cap B) = \frac{n(A \cap B)}{n(B)} \rightarrow \text{①}$$

For the conditional Probability i.e A/B the happening of ' A ' under the condition B has happened, the favourable outcomes (sample points) must be out of the Sample points of B i.e for the event A/B , the Sample Space is B .

$$\Rightarrow \boxed{P(B/A) = \frac{n(B \cap A)}{n(A)}}$$

from the Equation (1)

$$P(A \cap B) = \frac{n(A)}{n(S)} \times \frac{n(B \cap A)}{n(A)}$$

$$\boxed{P(A \cap B) = P(A) \times P(B/A)}$$

Similarly $P(B \cap A) = \frac{n(B)}{n(S)} \times \frac{n(B \cap A)}{n(B)}$

$$\boxed{P(B \cap A) = P(B) \times P(A/B)}$$

Hence, the Multiplication Theorem is Proved.

The events are said to be independent of each other if the happening of any one of them is not effected by and does not effect the happening of any one of the others. If the A & B are independent events, so that the probability of occurrence or non-occurrence of 'A', is not effected by the occurrence or non-occurrence of B then we have

$$\boxed{\begin{aligned} P(A/B) &= P(A) \\ P(B/A) &= P(B) \end{aligned}}$$

So, the multiplication theorem for two independent events / Theorem A & B

$$\boxed{\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(B \cap A) &= P(B) \cdot P(A) \end{aligned}}$$

BAYE'S THEOREM (RULE OF INVERSE PROBABILITY)

One of the important Application of the conditional probability is computations unknown probabilities on the bases of the information supplied by the experiment (or) past records. For example, Suppose an event has occurred through one of the various mutually disjoint events (mutually exclusive). Then the conditional probability that it has occurred due to a particular event reasons called Inverse Probability (or) posterior. These probabilities are computed by BAYE'S rule named after the British mathematician "Thomas Baye" who propounded it in 1763.

Theorem :-

If an event 'A' can occur only in conjunction with one of the 'n' mutually exclusive and exhaustive events $E_1, E_2, E_3, \dots, E_n$ and if A actually happens then the probability that it was preceded by a particular event

$E_i (i=1, 2, 3, \dots, n)$ is given by

$$P(E_i/A) = \frac{P(A \cap E_i)}{\sum_{i=1}^n P(E_i) \cdot P(A/E_i)} = \frac{P(E_i) P(A/E_i)}{\sum_{i=1}^n P(E_i) P(A/E_i)}$$

Proof:- Since the Event 'A' can occur in combination with any of the mutually exclusive and exhaustive events $E_1, E_2, E_3, \dots, E_n$ we have

$$A = (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup \dots \cup (A \cap E_n)$$

Where $A \cap E_1, A \cap E_2, \dots, A \cap E_n$ being the subsets of the mutually exclusive events $E_1, E_2, E_3, \dots, E_n$ all are disjoint i.e mutually exclusive events, Hence by Addition Theorem of probability

$$\begin{aligned}
 P(A) &= P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_n) \\
 &= P(E_1) P(A/E_1) + P(E_2) P(A/E_2) + \dots + P(E_n) P(A/E_n) \\
 &= \sum_{i=1}^n P(E_i) P(A/E_i)
 \end{aligned}$$

And for any particular event E_i , the conditional probability $P(E_i/A)$ is given by

$$P(E_i \cap A) = P(A) \cdot P(E_i/A)$$

$$P(E_i/A) = \frac{P(E_i \cap A)}{P(A)}$$

$$P(E_i/A) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^n P(E_i) \cdot P(A/E_i)}$$

Hence Proved.

Random - you can't predict the outcome

29/11/17

PROBABILITY DISTRIBUTION:-

A variable is something that can take a number of different values. A random variable is one that can take a no. of different values, but it is not possible to know which value it takes until the same experiment is performed.

In other words a random variable in a description of numerical values that the outcomes from an experiment can take random variables are of two types.

1. Discrete.

2. Continuous.

A random variable is said to be a "discrete" random variable if its possible outcomes are "finite" or "countable".

If a random variable represents measurements on a continuous scale, it is called "continuous" random variable.

The Probability distribution of 3 types

discrete $\left\{ \begin{array}{l} 1. \text{ Binomial Distribution.} \\ 2. \text{ Poisson Distribution.} \end{array} \right.$
continuous — 3. Normal Distribution.

1. BINOMIAL DISTRIBUTION:-

It is also known as Bernoulli Distribution in the name of the Swiss mathematician "Jacob Perlli".

Consider a fair coin tossed once. There can be two possibilities either head (success) or Tail (failure). Suppose if the probability of success is 'p' and the probability of failure is 'q' then " $p+q$ ".

$p+q$ = Sum of the probabilities

Consider two coins tossed, then there are four possibilities

- i) Both coins fall head i.e. (p^2)
- ii) first coin falling head & second coin falling Tail i.e. (pq)
- iii) First coin falling tail & second coin falling head i.e. qp
- iv) Both coins fall Tail i.e. q^2

Thus the probabilities of zero, one and two successes is given by q^2 , $2pq$, p^2 respectively. Which is obtained from the Expansion of the Binomial.

$$(q+p)^2 = q^2 + p^2 + 2pq$$

Similarly, if 3 coins are tossed simultaneously, the probabilities of zero, one, two and three success is given by the term

q^3 , $3pq^2$, $3qp^2$, p^3 which can be expanded using the binomial

$$(q+p)^3 = q^3 + p^3 + 3p^2q + 3qp^2$$

Summarizing the above results, we have

i) One coin or one event $q+p = p+q$

ii) Two coins or Two events $(q+p)^2 = p^2 + q^2 + 2pq$

iii) Three coins or Three events $(p+q)^3 = p^3 + q^3 + 3p^2q + 3p^2q$

Hence for 'n' coins or 'n' events

$$(q+p)^n = q^n n p^{n-1} + \frac{n(n-1)}{2!} q^{n-2} p^2 + \frac{n(n-1)(n-2)}{3!} q^{n-3} p^3 + \dots + p^n$$

This is known as the "Binomial Distribution". The Binomial Distribution $(q+p)^n$ in general terms is written as $nC_r q^{n-r} p^r$

Where $nC_r = \frac{n!}{r!(n-r)!}$ (or) $\frac{n!}{r! (n-r)!}$

POISSON DISTRIBUTION:-

The Poisson Distribution was developed by a French Mathematician Poisson. The Poisson distribution is used in "Quality control statistics" to count the number of defective items and in many applications.

The characteristics of the Poisson distribution are

- i) It is described as a "Discrete Random Variable".
- ii) The events occur independently that is the occurrence of one event is not influenced by the occurrence of other event.
- iii) Theoretically, there is no upper limit with the number of occurrences of an event during a specified time period.
- iv) The probability of single occurrence of an event within a specific time period is proportional to the length of the time period.
- v) In an extremely small portion of the time period, the probability of two or more occurrences of an event is negligible.

The Poisson formula:-

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Where $p(x)$ = probability of 'x' occurrences

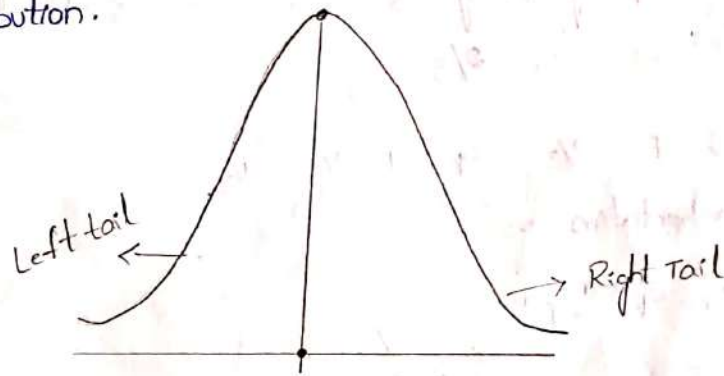
λ = lambda that is the mean number of occurrences for interval of time raised to the power of 'x'.

$e^{-\lambda}$ = 2.718281, being the base of natural logarithms raised to negative lambda.

$x!$ = x factorial.

NORMAL DISTRIBUTION:-

The Normal Distribution is an important continuous probability distribution which is also called "Gaussian Distribution" after the name of a mathematician "Karl Gauss". This normal distribution is widely used in statistics. The following graph represents the normal distribution.



The following are the characteristics of normal distribution

- i) The curve is "Bell shaped", that is it will have the same slope and shape on the either side of the vertical line from the mean.
- ii) It has a single peak, that uni model.
- iii) The mean is located at the centre of the distribution.
- iv) The distribution is symmetrical.
- v) The two tails of the distribution extend indefinitely but never touch the horizontal axis.
- vi) Since the normal curve is symmetrical the lower and upper "Quartiles" are equidistance from the median that is

$$Q_3 - \text{Median} = \text{Median} - Q_1$$

vii) The mean, Median and mode have the same value.

$$\text{Mean} = \text{Median} = \text{Mode}$$

viii) The units of - for the standard Normal Distribution curve is denoted by "Z" which is called Z values or Z scores.

$$Z = \frac{x - \mu}{\sigma}$$

Where

Z = Normal Deviate.

x = Variable under study.

μ = Mean.

σ = Standard Deviation.

1/12/17

i) Find the chance of getting 3 successes in 5 trials when the chance of getting a success is $\frac{2}{3}$.

Sol

$$r = 3, n = 5; p = \frac{2}{3}, q = 1 - \frac{2}{3} = \frac{1}{3}$$

The Binomial representation is

$$(p+q)^n = {}^n C_r \cdot q^{n-r} \cdot p^r$$

$$= \frac{5!}{3!(5-3)!} \cdot \left(\frac{1}{3}\right)^{5-3} \cdot \left(\frac{2}{3}\right)^3$$

$$= 10 \left(\frac{1}{3}\right)^2 \left(\frac{8}{27}\right)$$

$$= 10 \left(\frac{1}{9}\right) \left(\frac{8}{27}\right)$$

$$= 10 \times \frac{8}{243} = 0.329$$

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

$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 (5-3)!}$$

$$= \frac{120}{12}$$

$$= 10$$

* * * For a Binomial Distribution the mean is 4. The variance is 2. Find the probability of getting i) Atleast two successes (minimum)

* * * ii) utmost 2 successes

Sol In binomial distribution with 'n trials' and 'p' Successes are considered the mean and variance are as follows.

$$\text{Mean} = n \cdot p$$

$$\text{Variance} = npq$$

$$\text{S.D} = \sqrt{npq}$$

$$\text{Given mean} = 4$$

$$\text{Variance} = 2$$

$$\therefore np = 4$$

$$npq = 2 \Rightarrow 4q = 2$$

$$q = 2/4$$

$$q = 1/2$$

$$0! = 1$$

$$nC_0 = 1$$

$$nC_1 = n$$

$$nC_n = 1$$

$$\rightarrow \text{We have } p + q = 1$$

$$p + 1/2 = 1$$

$$p = 1 - 1/2 = 1/2 \quad \therefore p = 1/2$$

$$\rightarrow \text{Substitute } p = 1/2 \text{ in mean } np = 4$$

$$n(1/2) = 4$$

$$n = 4 \times 2 = 8 \quad \therefore n = 8$$

1) Atleast two (2) Successes:-

The probability of getting atleast two successes means, the probabilities of 2, 3, 4, 5, 6, 7 & 8 Successes.

For 2 Successes:- $r = 2$

$$(p+q)^n = nC_r \cdot p^r \cdot q^{n-r} = 8C_2 \cdot \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{8-2}$$

$$= \frac{8!}{2!(8-2)!} \left(\frac{1}{4}\right) \left(\frac{1}{2}\right)^6$$

$$= \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1) (6 \times 5 \times 4 \times 3 \times 2 \times 1)} \left(\frac{1}{4}\right) \left(\frac{1}{64}\right)$$

$$= \frac{28}{256}$$

$$r \text{ at } 2 = 0.1094$$

for $r = 3$ Successes:-

$$(p+q)^n = nC_r \cdot p^r \cdot q^{n-r} = 8C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{8-3}$$

$$= \frac{8!}{3!(8-3)!} \left(\frac{1}{8}\right) \left(\frac{1}{32}\right)$$

$$= \frac{\cancel{8} \times \cancel{7} \times \cancel{6} \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) (5 \times 4 \times 3 \times 2 \times 1)} \cdot \frac{1}{8} \times \frac{1}{32}$$

$$= \frac{7}{32} = 0.22$$

$$\boxed{r_3 = 0.2187}$$

for $r=4$ Successes

$$(p+q)^n = nC_r p^r q^{n-r}$$

$$= {}^8C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{8-4}$$

$$= \frac{8!}{4!(8-4)!} \left(\frac{1}{16}\right) \left(\frac{1}{16}\right)$$

$$= \frac{\cancel{8} \times \cancel{7} \times \cancel{6} \times \cancel{5} \times 4 \times 3 \times 2 \times 1}{4! (4 \times 3 \times 2 \times 1)} \left(\frac{1}{16}\right) \left(\frac{1}{16}\right)$$

$$= \frac{35}{128} = 0.2734$$

$$\boxed{r_4 = 0.2734}$$

for $r=5$

$$(p+q)^n = nC_r p^r q^{n-r}$$

$$= {}^8C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{8-5}$$

$$= \frac{\cancel{8} \times \cancel{7} \times \cancel{6} \times \cancel{5} \times 4 \times 3 \times 2 \times 1}{5! (8-5)!} \times \left(\frac{1}{32}\right) \left(\frac{1}{8}\right)$$

$$= \frac{7}{32}$$

$$\boxed{r_5 = 0.2187}$$

for $r=6$:-

$$nC_r p^r q^{n-r}$$

$$= {}^8C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^{8-6}$$

$$= \frac{\cancel{8} \times \cancel{7} \times \cancel{6} \times 5 \times 4 \times 3 \times 2 \times 1}{6! (2 \times 1)} \left(\frac{1}{64}\right) \left(\frac{1}{4}\right)$$

$$= \frac{7}{64} = 0.1094$$

$$\boxed{r_6 = 0.1094}$$

for $r=7$:- $nCr p^r q^{n-r} = {}^8C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^1$

$$= \frac{8 \times 7!}{7! 1} \times \frac{1}{128} \times \frac{1}{2}$$

$$= \frac{1}{32} = 0.0312$$

$$r_7 = 0.0312$$

for $r=8$:- $nCr p^r q^{n-r} = {}^8C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^{8-8}$

$$= \frac{8!}{8!(8-8)!} \cdot \frac{1}{256} (1)$$

$$= \frac{1}{256} = 0.0039$$

$$r_8 = 0.0039$$

for Total Atleast 2 success = $0.1094 + 0.2187 + 0.2734 + 0.2187 + 0.1094 + 0.0312 + 0.0039$

$$\text{Atleast 2} = 0.9648$$

Utmost 2 Success:- Mean of 0, 1 & 2 Successes

$$r=0 \Rightarrow {}^8C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{8-0} = \frac{8!}{0! 8!} (1) \left(\frac{1}{256}\right) = 0.0039$$

$$r=1 \Rightarrow {}^8C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{8-1} = \frac{8 \times 7!}{1! \times 7!} \left(\frac{1}{2}\right) \left(\frac{1}{128}\right) = \frac{1}{32} = 0.0312$$

$$r=2 \Rightarrow 0.1094$$

$$\text{Utmost 2 success} = 0.0039 + 0.0312 + 0.1094$$

$$= 0.1445$$

Wednesday
6/12/17

Fitting a Binomial Distribution:-

The following are the steps in fitting a Binomial Distribution

1. Given Some Information, find the values of p & q , when one value is given, the second value can be obtained by subtracting the first value from 'One' (1).

2. Find the Binomial $(p+q)^n$, here it is important to note that the power of 'n' will be one less than the number of terms in the expanded Binomial. For example - when $n=5$, there will be 6 terms.

3. Multiply each of the expanded binomial terms by the total frequency 'N' so that the expected frequency in each category can be obtained.

1) fit a Binomial Distribution to the following data

x	0	1	2	3	4
f	28	62	46	10	4

Sol

x	f	fx
0	28	0
1	62	62
2	46	92
3	10	30
4	4	16
	<u>150</u>	<u>200</u>

$$\begin{aligned}\text{Mean} &= \frac{\sum fx}{N} \\ &= \frac{200}{150} \\ &= 1.33\end{aligned}$$

In binomial distribution Mean $\boxed{np = 1.33}$

$$4p = 1.33$$

$$p = \frac{1.33}{4} \Rightarrow 0.33$$

$$\Rightarrow p + q = 1$$

$$\Rightarrow 0.33 + q = 1$$

$$\Rightarrow q = 0.67$$

$$Y \quad f(x) = N \cdot {}^n C_r (p)^r (q)^{n-r}$$

Theoretical frequency

0	$150 \times {}^4 C_0 (0.33)^0 (0.67)^{4-0}$	30.23
1	$150 \times {}^4 C_1 (0.33)^1 (0.67)^{4-1}$	59.55
2	$150 \times {}^4 C_2 (0.33)^2 (0.67)^{4-2}$	43.99
3	$150 \times {}^4 C_3 (0.33)^3 (0.67)^{4-3}$	14.45
4	$150 \times {}^4 C_4 (0.33)^4 (0.67)^{4-4}$	1.78
		<u>150</u>

At $r=0$

$$\begin{aligned} & N \cdot nCr \cdot p^r \cdot q^{n-r} \\ \Rightarrow & 150 \times {}^4C_0 (0.33)^0 (0.67)^{4-0} \\ & = 150 \times 1 \times 1 \times 0.2015 \\ & = 30.23 \end{aligned}$$

At $r=2$

$$\begin{aligned} & N \cdot nCr \cdot p^r \cdot q^{n-r} \\ & = 150 \times {}^4C_2 (0.33)^2 (0.67)^{4-2} \\ & = 150 \times \frac{4!}{2!(4-2)!} \times 0.1089 \times 0.4489 \\ & = 150 \times \frac{4 \times 3 \times 2 \times 1}{(2 \times 1)(2 \times 1)} \times 0.1089 \times 0.4489 \\ & = 150 \times 6 \times 0.1089 \times 0.4489 \\ & = 43.9967 \end{aligned}$$

At $r=4$

$$\begin{aligned} & N \cdot nCr \cdot p^r \cdot q^{n-r} \\ & = 150 \times {}^4C_4 (0.33)^4 (0.67)^{4-4} \\ & = 150 \times \frac{4!}{4!0!} \times 0.0119 \times 1 \\ & = 150 \times 0.0119 \\ & = 1.7789 \end{aligned}$$

At $r=1$

$$\begin{aligned} & N \cdot nCr \cdot p^r \cdot q^{n-r} \\ & = 150 \times {}^4C_1 (0.33)^1 (0.67)^{4-1} \\ & = 150 \times \frac{4!}{1!(4-1)!} \times 0.33 \times 0.3008 \\ & = 150 \times 4 \times 0.33 \times 0.3008 \\ & = 59.55 \end{aligned}$$

At $r=3$

$$\begin{aligned} & N \cdot nCr \cdot p^r \cdot q^{n-r} \\ & = 150 \times {}^4C_3 (0.33)^3 (0.67)^{4-3} \\ & = 150 \times \frac{4!}{3!1!} \times 0.0359 \times 0.67 \\ & = 150 \times 4 \times 0.0359 \times 0.67 \\ & = 14.43 \end{aligned}$$

Poisson Distribution Problems:-

1) If the proportion of the defective items is $p=0.01$, A random sample of 100 items is selected. What is the probability that there are two defective items.

Sol In poisson Distribution

$\lambda = \text{mean}$

$\Rightarrow \lambda = n \cdot p$

$$\Rightarrow \lambda = 100 \times 0.01$$

$$\boxed{\text{mean } \lambda = 1}$$

Poisson Distribution is

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

at $x=2$

$$P(2) = \frac{e^{-1} 1^2}{2!}$$

$$= \frac{1}{2e}$$

$$= \frac{1}{2 \times 2.718281}$$

$$= \frac{1}{5.4366}$$

$$= 0.1839$$

2) The probability of dialing a wrong number is 0.05. What is the probability of dialing exactly 3 wrong numbers in 100 dials.

$$n=100; p=0.05$$

$$\lambda = np$$

$$\lambda = 100 \times 0.05$$

$$\boxed{\text{Mean } \lambda = 5}$$

at $x=3$

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$P(3) = \frac{e^{-5} 5^3}{3!}$$

$$= \frac{5^3}{3! e^5}$$

$$= \frac{125}{6 \times 148.4132} = \frac{125}{890.4790}$$

$$= 0.1404$$

7/12/17

3) fit a poisson distribution to the following data

Deaths -	0	1	2	3	4
frequency -	109	65	22	3	1

Sol

x	f	fx
0	109	0
1	65	65
2	22	44
3	3	9
4	1	4
	<u>200</u>	<u>122</u>

$$N = 200$$

$$\sum fx = 122$$

$$\begin{aligned} \text{Mean} &= \frac{\sum fx}{N} \\ &= \frac{122}{200} \\ &= 0.61 \end{aligned}$$

In poisson Distribution mean is respectively ' λ '

$$\Rightarrow \lambda = n \cdot p$$

$$\Rightarrow 0.61 = 4p$$

$$\Rightarrow p = 0.1525$$

$$\therefore q = 1 - 0.1525$$

$$q = 0.8475$$

x	$N \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!}$
0	$200 \cdot \frac{(0.61)^0 \cdot e^{-0.61}}{0!}$
1	$200 \cdot \frac{(0.61)^1 \cdot e^{-0.61}}{1!}$
2	$200 \cdot \frac{(0.61)^2 \cdot e^{-0.61}}{2!}$
3	$200 \cdot \frac{(0.61)^3 \cdot e^{-0.61}}{3!}$
4	$200 \cdot \frac{(0.61)^4 \cdot e^{-0.61}}{4!}$

Theoretical frequency

$$108.67$$

$$66.29$$

$$20.22$$

$$4.11$$

$$0.637$$

$$1$$

$$e^{-0.61} = 0.5434$$

4) The following table shows the number of days in 50 day period during which x number of automobile accidents occurred in a city. fit a poisson distribution

x	0	1	2	3	4
f	21	18	7	3	1

Sol

x	f	fx
0	21	0
1	18	18
2	7	14
3	3	9
4	1	4
	<u>50</u>	<u>45</u>

$$N = 50$$

$$\Sigma fx = 45$$

$$\text{Mean} = \frac{\Sigma fx}{N}$$

$$= \frac{45}{50}$$

$$\lambda = 0.9$$

In poisson Distribution on mean is represented by ' λ '

$$\lambda = np$$

$$0.9 = 4p$$

$$p = 0.225$$

$$p + q = 1 \Rightarrow q = 1 - p$$

$$q = 1 - 0.225$$

$$q = 0.775$$

$$e^{-0.9} = 0.4066$$

x	$N \cdot \frac{e^{-\lambda} \lambda^x}{x!}$
0	$50 \cdot \frac{e^{-0.9} (0.9)^0}{0!}$
1	$50 \cdot \frac{e^{-0.9} (0.9)^1}{1!}$
2	$50 \cdot \frac{e^{-0.9} (0.9)^2}{2!}$
3	$50 \cdot \frac{e^{-0.9} (0.9)^3}{3!}$
4	$50 \cdot \frac{e^{-0.9} (0.9)^4}{4!}$

Theoretical frequency

$$20.3285$$

$$18.2970$$

$$8.2337$$

$$2.4701$$

$$0.5558$$

$$\underline{49.89}$$

5) The number of Arrivals of ships per day at a seaport are as given below

No. of Arrivals : 0 1 2 3 4 5 6 7 8 9 10

No. of Days : 20 50 80 83 64 34 18 8 3 1 1
can it be concluded that the no. of Arrivals follow a Poisson Distribution

Sol

x	f	fx
0	20	0
1	50	50
2	80	160
3	83	249
4	64	256
5	34	170
6	18	108
7	8	56
8	3	24
9	1	9
10	1	10
	<u>362</u>	<u>1092</u>

$$N = 362$$

$$\Sigma fx = 1092$$

$$\begin{aligned} \text{Mean} &= \frac{\Sigma fx}{N} \\ &= \frac{1092}{362} \\ &= 3.02 \end{aligned}$$

In poisson Distribution $\lambda = \text{mean}$

$$\text{Hence } \lambda = 3.02$$

$$\lambda = np$$

$$P + Q = 1 \quad 3.02 = 10 \cdot p$$

$$Q = 1 - P \quad P = 0.3021$$

$$Q = 1 - 0.3021$$

$$Q = 0.6980$$

x	$N \cdot \frac{e^{-\lambda} \lambda^x}{x!}$	Theoretical frequency
0	$362 \cdot \frac{(3.02)^0 e^{-3.02}}{0!} = 362 \times 1 \times 0.0488$	17.666
1	$362 \cdot \frac{(3.02)^1 e^{-3.02}}{1!} = 362 \times 3.02 \times 0.0488$	53.3514
2	$362 \cdot \frac{(3.02)^2 e^{-3.02}}{2!} = 362 \times \frac{9.1204 \times 0.0488}{2!}$	80.5587
3	$362 \cdot \frac{(3.02)^3 e^{-3.02}}{3!} = 362 \times \frac{27.5436 \times 0.0488}{6}$	81.0957
4	$362 \cdot \frac{(3.02)^4 e^{-3.02}}{4!} = 362 \times \frac{83.1817 \times 0.0488}{24}$	61.2273
5	$362 \cdot \frac{(3.02)^5 e^{-3.02}}{5!} = 362 \times \frac{251.2087 \times 0.0488}{5 \times 4 \times 3 \times 2 \times 1}$	36.9813
6	$362 \cdot \frac{(3.02)^6 e^{-3.02}}{6!} = 362 \times \frac{758.6503 \times 0.0488}{720}$	18.6139
7	$362 \cdot \frac{(3.02)^7 e^{-3.02}}{7!} = 362 \times \frac{2291.1240 \times 0.0488}{5040}$	8.0306
8	$362 \cdot \frac{(3.02)^8 e^{-3.02}}{8!} = 362 \times \frac{6919.1946 \times 0.0488}{40320}$	3.0315
9	$362 \cdot \frac{(3.02)^9 e^{-3.02}}{9!} = 362 \times \frac{20895.9676 \times 0.0488}{362880}$	1.0173
10	$362 \cdot \frac{(3.02)^{10} e^{-3.02}}{10!} = 362 \times \frac{63105.8222 \times 0.0488}{3628800}$	0.3072

$$e^{-3.02} = 0.0488$$

We observe that the Sum of the Expected frequencies (362) is equal to The Sum of the Theoretical frequency (361.89). Hence it can be concluded that the number of arrivals follow poisson Distribution.

Poisson Distribution as an Approximation to Binomial Distribution:-

-As the Binomial distribution involves tedious (heavy) calculations, Poisson distribution can be used in its place. However this holds good under certain conditions such as Number of trials is large in which the Binomial probability of success is small i.e. $n \geq 20$, $p \leq 0.05$.

The Poisson distribution will become a limiting form of Binomial distribution.

In Binomial Distribution

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$\text{mean} = np$$

Where $\lambda = np \Rightarrow$ $P(x) = \frac{(np)^x e^{-np}}{x!}$ - Limiting form of B.D

2010

- i) In a certain factory turning out Razor Blades, there is a small chance of $1/500$ for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective, one defective, two defectives and three defectives in a consignment of 10,000 packets.
- ii) Define Normal distribution. What are the properties of Normal curve. Describe the importance of Normal distribution.

Sol Probability of any blade to be defective = $\frac{1}{500} = 0.002$

$\Rightarrow$ We know that In Binomial Distribution Mean = np

$$\text{Mean} = np$$

$$= 10 \times 0.002$$

$$= 0.02$$

$$= 0.02$$

$$\lambda = 0.02$$

According to poisson distribution $\Rightarrow P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

we have

$$\text{No defective} \Rightarrow x=0 \Rightarrow P(0) = \frac{(0.02)^0 \cdot e^{-0.02}}{0!} \\ = 0.9802$$

$$\text{one defective} \Rightarrow x=1 \Rightarrow P(1) = \frac{(0.02)^1 \cdot e^{-0.02}}{1!} \\ = 0.0196$$

$$\text{Two defective} \Rightarrow x=2 \Rightarrow P(2) = \frac{(0.02)^2 \cdot e^{-0.02}}{2!} = \frac{0.0004 \times 0.9802}{2} \\ = 0.0002$$

$$\text{Three defective} \Rightarrow x=3 \Rightarrow P(3) = \frac{(0.02)^3 \cdot e^{-0.02}}{3!} \\ = 0$$

At $x=0$

For 10000 packets $\Rightarrow 10000 \times 0.9802 = 9802$ packets (No defective)

At $x=1$

for 10000 packets $\Rightarrow 10000 \times 0.0196 = 196$ packets (one defective)

At $x=2$

For 10000 packets $\Rightarrow 10000 \times 0.0002 = 2$ packets (Two defectives)

At $x=3$

For 10000 packets $\Rightarrow 10000 \times 0 = 0$ packets (Three defectives).

2014

3) Define binomial distribution. What are its characteristics.

It is known from the past experience that 80% of the students in a school do their homework. Find the probability that during a random check of 10 students

i) All have done their homework.

ii) At least most two have done their homework.

iii) At least one has not done their homework.

Sol Let $p = \text{probability of doing homework} = 80\% = \frac{80}{100} = 0.8$

Then $q = \text{probability of not doing homework} = 1 - 0.8 = 0.2$

Number of students $n = 10$

i) All have done their homework.

$$\text{Binomial Distribution} = {}^nC_r p^r q^{n-r}$$

$$n=10, r=10$$

$$= {}^{10}C_{10} (0.8)^{10} (0.2)^{10-10}$$

$$= 1 \times 0.8^{10} \times 1$$

$$= 0.1074$$

ii) Atleast most two have done their homework

$$r=2$$

$$\Rightarrow {}^nC_r p^r q^{n-r}$$

$$= {}^{10}C_2 (0.8)^2 (0.2)^{10-2}$$

$$= \frac{10!}{2!(10-2)!} \times 0.8^2 \times (0.2)^8$$

$$= \frac{10 \times 9 \times 8!}{2! \cdot 8!} \times 0.64 \times 0.0000025$$

$$\Rightarrow 0.000073$$

iii) Atleast one has not done their homework.

$$r = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

$$\text{At } r=1$$

$$\Rightarrow {}^{10}C_1 (0.8)^1 (0.2)^{10-1} = 10 \times 0.8 \times (0.2)^9$$

$$= 0.000004096$$

$$\text{At } r=2$$

$$\Rightarrow {}^{10}C_2 (0.8)^2 (0.2)^{10-2}$$

$$= \frac{10!}{2!(10-2)!} \times 0.64 \times 0.0000025$$

$$= \frac{10 \times 9 \times 8!}{2! \cdot 8!} \times 0.64 \times 0.0000025$$

$$= 0.000072$$

At r=3

$$\Rightarrow 10C_3 (0.8)^3 (0.2)^{10-3} \Rightarrow \frac{10 \times 9 \times 8 \times 7!}{(3 \times 2 \times 1) 7!} \times 0.512 \times 0.0000128$$

$$= 0.000786432$$

At r=4

$$\Rightarrow 10C_4 (0.8)^4 (0.2)^{10-4}$$

$$= 10C_4 \times 0.4096 \times (0.2)^6 \Rightarrow \frac{10 \times 9 \times 8 \times 7 \times 6!}{(4 \times 3 \times 2 \times 1) 6!} \times 0.4096 \times 0.000064$$

$$\Rightarrow 210 \times 0.4096 \times 0.000064$$

$$= 0.005505024$$

At r=5

$$\Rightarrow 10C_5 \times (0.8)^5 (0.2)^5 \Rightarrow 252 \times 0.32768 \times 0.00032$$

$$= 0.026424115$$

At r=6

$$\Rightarrow 10C_6 \times (0.8)^6 (0.2)^4 \Rightarrow 210 \times 0.262144 \times 0.0016$$

$$= 0.08808$$

At r=7

$$\Rightarrow 10C_7 \times (0.8)^7 (0.2)^3 \Rightarrow 120 \times 0.2097152 \times 0.008$$

$$= 120 \times 0.001678122 = 0.20132$$

At r=8

$$\Rightarrow 10C_8 \times (0.8)^8 \times (0.2)^2 = 45 \times 0.1678 \times 0.04 = 0.3020$$

At r=9

$$\Rightarrow 10C_9 \times (0.8)^9 \times (0.2)^1 = 10 \times 0.1342 \times 0.2 = 0.2684$$

At r=10

$$\Rightarrow 10C_{10} \times (0.8)^{10} \times (0.2)^0 = 1 \times 0.1074 \times 1 = 0.1074$$

Total

$$= 0.000004096 + 0.00007372 + 0.0007864 + 0.005505$$

$$+ 0.02642 + 0.08808 + 0.20132 + 0.3020 + 0.2684 + 0.1074$$

$$= 1.$$

NORMAL DISTRIBUTION

- 1) Given that a random variable x is distributed normally with a Mean of 70 and a variance of 36, find the Probability of
- i) 82 and greater ii) 75 and greater iii) Between 75 and 85

Sol

Given that Mean $\mu = 70$

Variance $\sigma^2 = 36$

$$\sqrt{\text{Variance}} = \text{S.D}$$

$$\Rightarrow \sqrt{36} = 6$$

i) At $x = 82$

$$\Rightarrow Z = \frac{82 - 70}{6}$$
$$= \frac{12}{6} = 2$$

$$\Rightarrow \boxed{Z = 2}$$

$$= 0.5 - 0.4772$$

$$= 0.0228 \times 100$$

$$= 2.28\%$$

ii) At $x = 75$

$$\Rightarrow Z = \frac{x - \mu}{\sigma}$$
$$= \frac{75 - 70}{6}$$
$$= \frac{5}{6} = 0.83$$

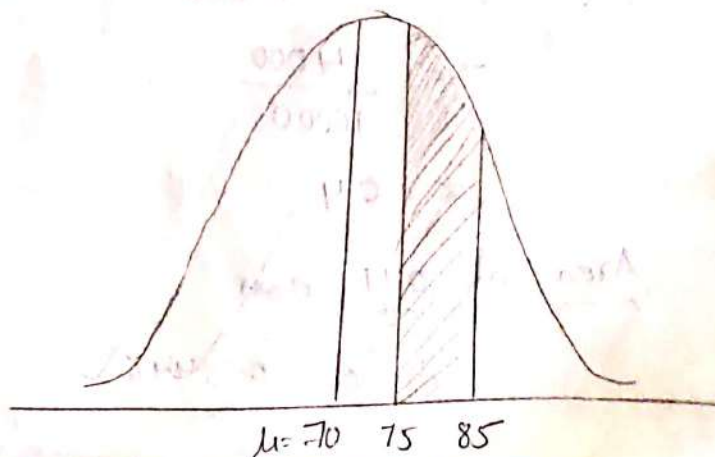
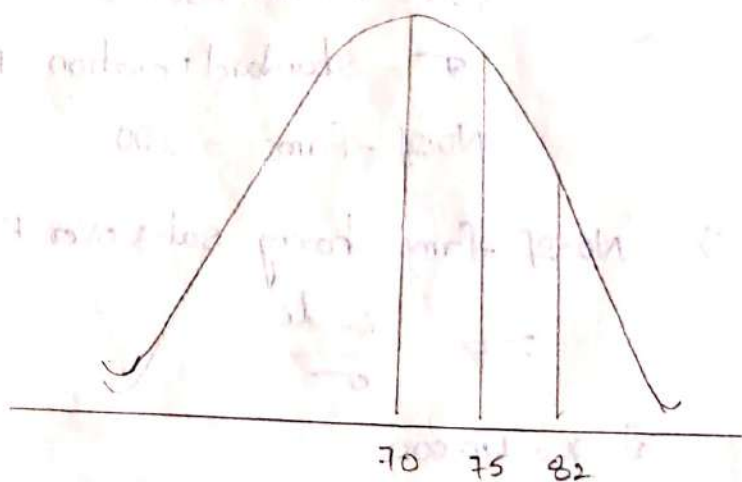
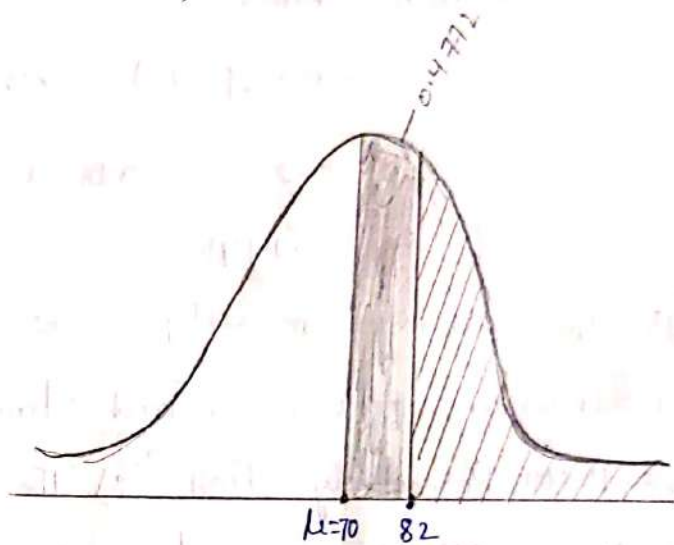
$$\text{Area} = 0.2033 \times 100$$

$$= 20.33\%$$

iii) Between 75 & 85

$$\text{At } 75 \quad Z = 0.83$$

$$\text{At } 85 \Rightarrow Z = \frac{x - \mu}{\sigma}$$
$$= \frac{85 - 70}{6}$$
$$= \frac{15}{6} = 2.5$$



$$\text{At } z = 0.83$$

$$\text{Area} = 0.2033$$

$$\text{At } z = 2.5$$

$$\text{Area} = 0.0062$$

iii) Required Area

$$= \text{Area of } (75) - \text{Area of } (85)$$

$$= 0.2033 - 0.0062$$

$$= 0.1971$$

*2) The Average monthly sales of 5000 firms are Normally distributed. Its mean and standard deviation are Rs. 36,000 and Rs. 10,000 respectively. Find i) The Number of firms having sales over Rs. 40,000. ii) The Percentage of firms having sales between Rs. 38,500 and Rs. 41,000 iii) The number of firms having sales between Rs. 30,000 and Rs. 40,000

Sol

Given that

$$\mu = \text{Mean} = \text{Rs. } 36,000$$

$$\sigma = \text{Standard Deviation} = \text{Rs. } 10,000$$

$$\text{No. of firms} = 5000$$

i) No. of firms having sales over Rs. 40,000

$$z = \frac{x - \mu}{\sigma}$$

$$\text{at } x = 40,000$$

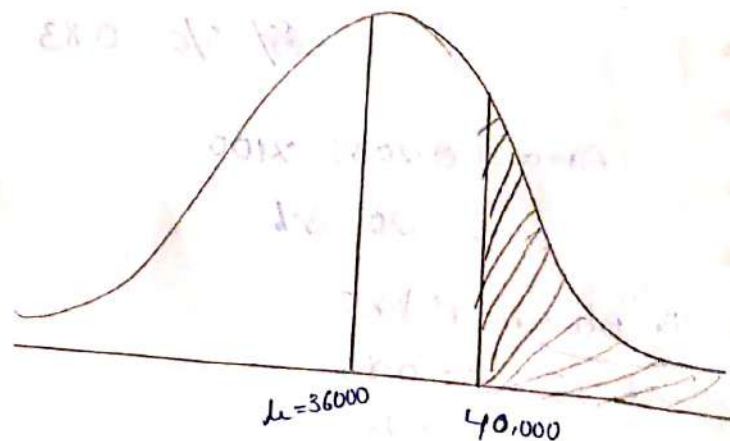
$$\therefore z = \frac{40,000 - 36,000}{10,000}$$

$$z = \frac{4,000}{10,000}$$

$$z = 0.4$$

Area at 0.4 corr

$$\text{Area} = 0.3446$$



$$\text{Number of firms} = 0.3446 \times 5000$$

$$= 1723 \text{ firms having Above Rs. 40,000 Sales.}$$

ii) Percentage of sales between Rs. 38,500 and 41,000

$$\text{At } x = 38500$$

$$\Rightarrow z = \frac{x - \mu}{\sigma}$$

$$= \frac{38500 - 36000}{10,000}$$

$$z = \frac{2500}{10000}$$

$$z = 0.25$$

At $z = 0.25$ corresponding Area is

$$\text{Area} = 0.4013$$

$$\text{At } x = 41,000$$

$$\Rightarrow z = \frac{x - \mu}{\sigma}$$

$$z = \frac{41000 - 36000}{10000}$$

$$= \frac{5000}{10000}$$

$$z = 0.5$$

$$\text{Corresponding Area} = 0.3085$$

Required Area

$$\Rightarrow 0.4013 - 0.3085$$

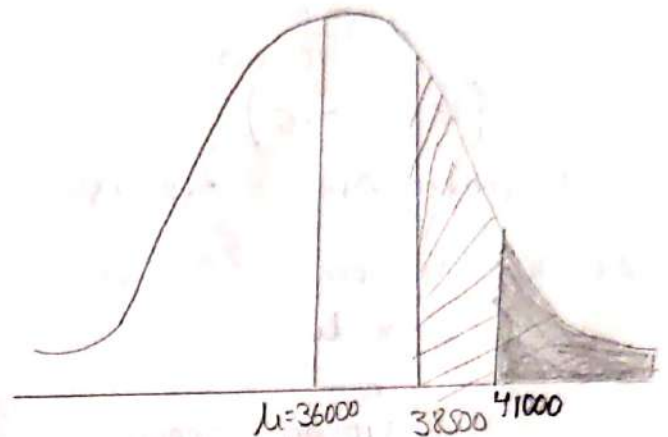
$$= 0.0928$$

$$\text{No. of firms} = 0.0928 \times 5000$$

$$= 464$$

Percentage of firms

$$\frac{464}{5000} \times 100 = 9.28$$



iii) between 30000 and 40000

At $x = 30000$

$$z = \frac{x - \mu}{\sigma}$$
$$= \frac{30000 - 36000}{10000}$$
$$= \frac{-6000}{10000}$$

$$z = -0.6$$

Corresponding Area = 0.2743

At $x = 40000$

$$z = \frac{x - \mu}{\sigma}$$
$$= \frac{40000 - 36000}{10000}$$
$$= \frac{4000}{10000}$$

$$z = 0.4$$

Corresponding Area = 0.3446

Required Area

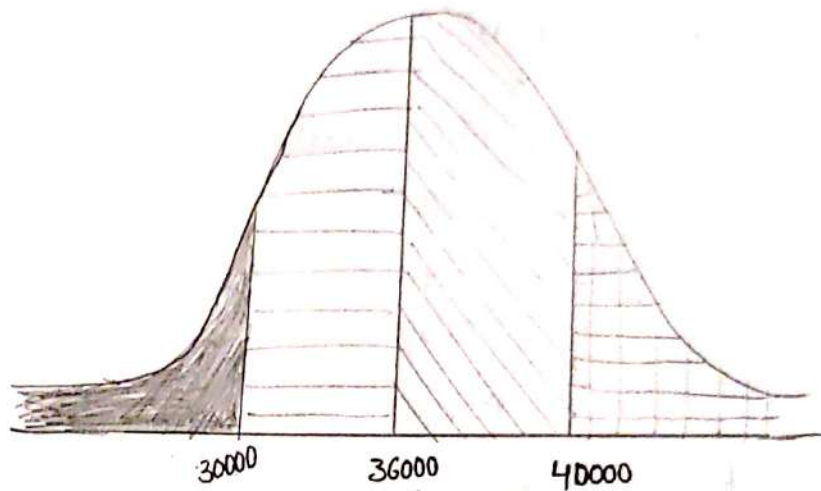
$$= 0.5 - 0.2743 + (0.5 - 0.3446)$$

$$= 0.2257 + 0.1554$$

$$= 0.3811$$

Number of firms = 0.3811×5000

$$= 1905.5$$



2015, 2016

3) The Average daily sales of 500 branches is Rs. 1,50,000 and the standard deviation is Rs. 15,000. Assuming the distribution to be Normal, Indicate how many branches have the sales between

i) Rs. 1,20,000 and Rs. 1,45,000

ii) Rs. 1,40,000 and Rs. 1,65,000

iii) more than Rs. 1,65,000

Sol Given that

$$\text{Mean} = \mu = 1,50,000$$

$$\text{S.D} = \sigma = 15,000$$

$$\text{No. of branches} = 500$$

$\therefore$ Rs. 1,20,000 and Rs. 1,45,000

At $x = 1,20,000$

$$z = \frac{x - \mu}{\sigma} = \frac{1,20,000 - 1,50,000}{15,000}$$
$$= \frac{-30,000}{15,000}$$

$$z = -2$$

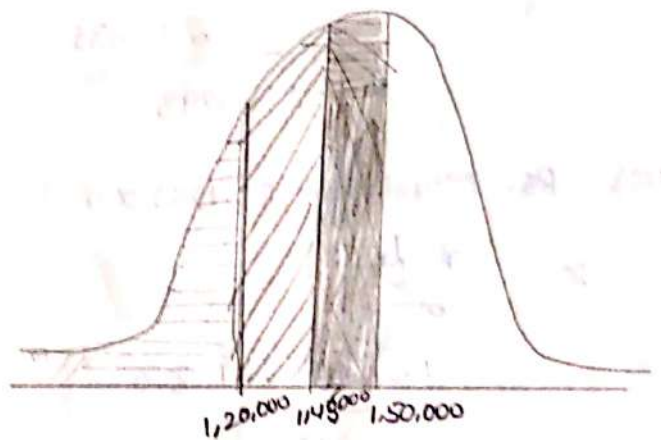
corresponding Area is 0.4772

At $x = 1,45,000$

$$z = \frac{x - \mu}{\sigma} = \frac{1,45,000 - 1,50,000}{15,000}$$
$$= \frac{-5,000}{15,000}$$

$$z = -0.33$$

corresponding Area is 0.1293



$$\therefore \text{Required Area} = 0.4772 + 0.1293$$
$$= 0.3479$$

$$\text{No. of Branches} = 0.3479 \times 500$$

$$= 173.95 \approx 174$$

$\therefore$ Rs. 1,40,000 and Rs. 1,65,000

At $x = 1,40,000$

$$z = \frac{1,40,000 - 1,50,000}{15,000} = \frac{-10,000}{15,000} = -0.66$$

At $x = 1,65,000$

Corresponding Area = 0.2454

At $x = 1,65,000$

$$Z = \frac{1,65,000 - 1,50,000}{15000}$$

$$= \frac{15000}{15000} = 1$$

Corresponding Area = 0.3413

Required Area is

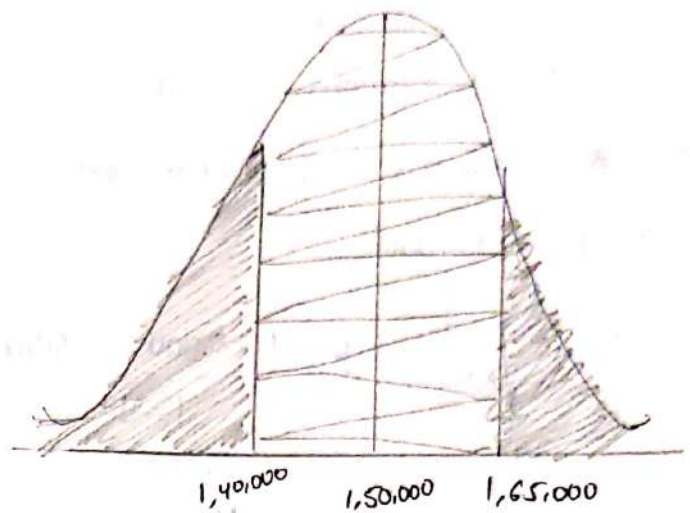
$$= 0.2454 + 0.3413$$

$$= 0.5867$$

No. of Branches = 0.5867×500

$$= 293.35$$

$$\approx 293$$

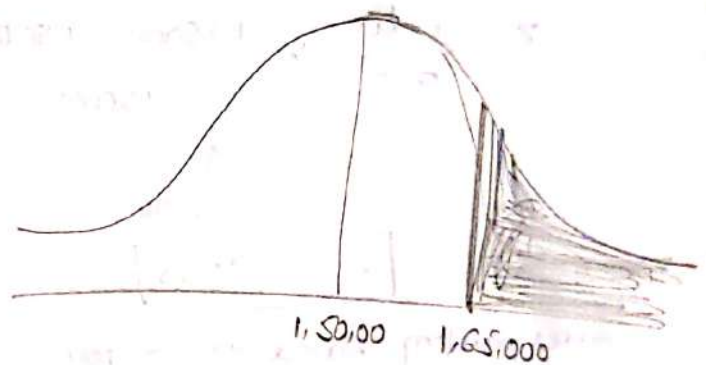


iii) Rs. More than Rs. 1,65,000

$$Z = \frac{x - \mu}{\sigma}$$

$$= \frac{1,65,000 - 1,50,000}{15000}$$

$$= \frac{15000}{15000} = 1$$



Required Area is $0.5 - 0.3413 = 0.1587$

No. of Branches = 0.3413×500 0.1587×100

$$= 170.65$$

$$\approx 171$$

$$\approx 79$$

4) AS a result of a test on 20,000 electric bulbs manufactured by a company, it was found that the life time of bulbs was normally distributed with an Average life of 2040 hrs and the standard deviation of 160 hrs. On the basis of this information, estimate the no. of bulbs that are expected to burn for

i) more than 2150 hrs. ii) less than 1960 hrs

Sol Given that

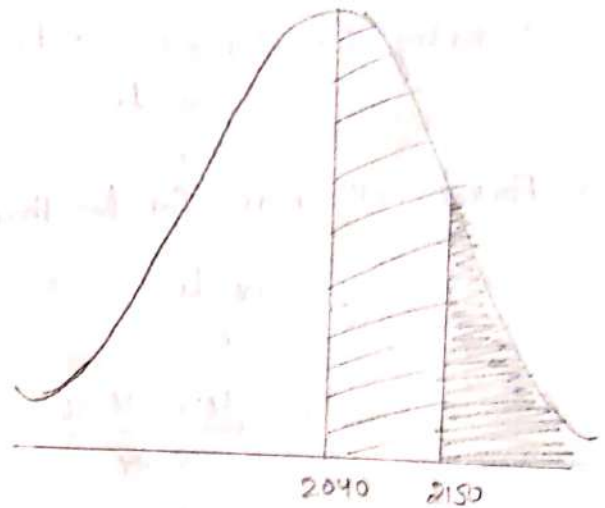
Mean $\mu = 2040$ hrs

S.D $\sigma = 60$ hrs

Total bulbs = 20,000

i) more than 2150 hrs

$$Z = \frac{x - \mu}{\sigma} = \frac{2150 - 2040}{60} \\ = \frac{110}{60} \\ = 1.83$$



Corresponding Area = ~~0.4664~~ 0.336

$$\text{No. of bulbs} = 0.336 \times 20000 \\ = 672 \text{ bulbs}$$

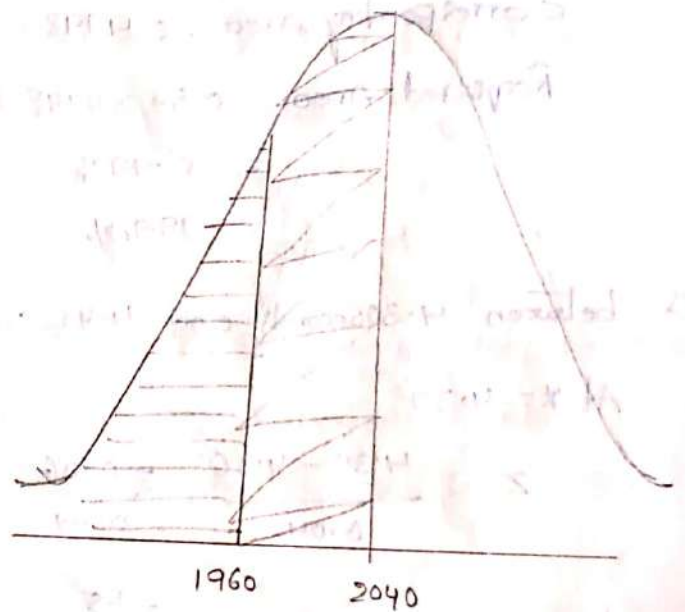
ii) less than 1960 hrs

$x = 1960$

$$Z = \frac{x - \mu}{\sigma} = \frac{1960 - 2040}{60} \\ = \frac{-80}{60} \\ = -1.33$$

Corresponding Area = 0.0918

$$\text{No. of bulbs} = 0.0918 \times 20000 \\ = 1836 \text{ bulbs}$$



Saturday
9/12/17

2015

- 5) i) Define poisson Distribution. What are its characteristics
- ii) The burning Time of an Experimental Rocket is a Random Variable which has a normal distribution with a mean of 4.36 sec and standard Deviation of 0.04 sec. What are the Probabilities that this kind of Rocket will burn for less than 4.5 sec, between 4.30 and 4.42 sec.

Sol

Given that

$$\text{Mean} = \mu = 4.36 \text{ seconds}$$

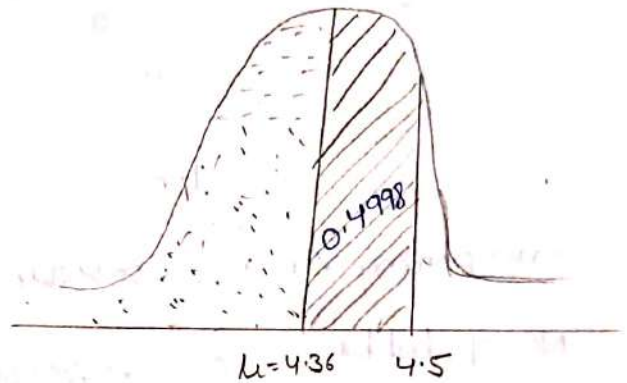
$$\text{S.D} = \sigma = 0.04 \text{ seconds}$$

According to Normal Distribution

$$z = \frac{x - \mu}{\sigma}$$

i) Rocket will burn for less than 4.5 seconds.

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{4.5 - 4.36}{0.04} \\ &= \frac{0.14}{0.04} \\ &= 3.5 \end{aligned}$$



At $z = 3.5$

Corresponding Area = 0.4998

$$\begin{aligned} \text{Required Area} &= 0.5 + 0.4998 \\ &= 0.9998 \\ &= 99.98\% \end{aligned}$$

ii) between 4.30 seconds and 4.42 seconds

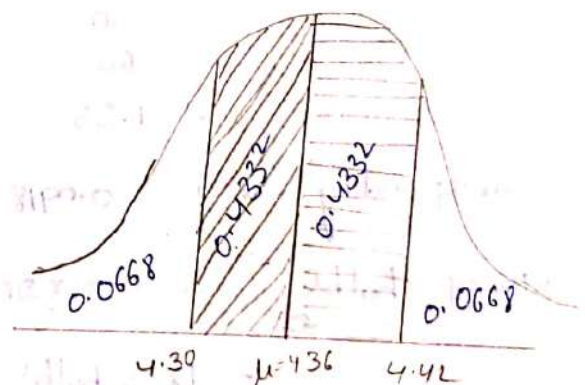
At $x = 4.30$

$$\begin{aligned} z &= \frac{4.30 - 4.36}{0.04} = \frac{-0.06}{0.04} \\ &= -1.5 \end{aligned}$$

At $z = -1.5$

Corresponding Area ^{from T-8} = 0.4332

$$\begin{aligned} \text{Corresponding Area from 761 page Table} &= 0.0668 \\ &= 10.5 - 0.0668 \\ &= 0.4332 \end{aligned}$$



At $x = 4.42$

$$z = \frac{4.42 - 4.36}{0.04} = \frac{0.06}{0.04} = 1.5$$

At $z = 1.5$

Corresponding Area = 0.4332

$$\text{Area} = 0.4332 + 0.4332$$

$$= 0.8664$$

$$= 86.64\%$$

2016

Q) Explain probability

Ans) Marginal probability, conditional probability and joint probabilities with examples.

Q) If 60% of the voters in a constituency favour a particular candidate, find the probability that in a sample of 400 voters, more than 250 voters would favour the candidate. Use Normal Approximation to Binomial.

Sol

Normal Approximation to Binomials -

When we come across a similar situation repeatedly and are interested in success (one of the two outcomes), then the total number of successes may follow a Binomial Distribution. A Binomial Distribution must satisfy certain conditions such as the probability of success must remain the same during the experiment and the outcome of an event must not be influenced by the outcome of any successive events.

When the probability of success is not very close to zero or one and the number of trials is large, then the normal distribution can be a good approximation to the Binomial distribution.

Q) In Binomial Distribution mean = np

$$\text{Given } n = 400$$

Given that 60% of the voters favour a particular candidate = $p = 60\%$
 $= 0.6$

$$\text{Now } \Rightarrow \text{Mean} = np$$

$$= 400 \times 0.6$$

$$= 240$$

In Binomial Distribution Variance = npq

$$q = 1 - p$$

$$q = 1 - 0.6$$

$$q = 0.4$$

$$\text{Variance } npq = 400 \times 0.6 \times 0.4 = 96$$

$$\text{Standard Deviation } \sigma = \sqrt{npq} = \sqrt{96} = 9.76$$

ay more than 250 voters

In Normal Distribution

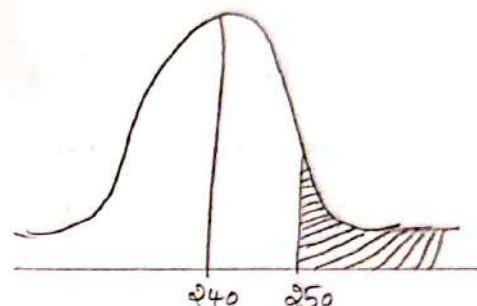
$$\text{Mean} = \mu = 240$$

$$\text{S.D} = \sigma = 9.79$$

$$Z = \frac{x - \mu}{\sigma}$$

$$= \frac{250 - 240}{9.79}$$

$$= \frac{10}{9.79} = 1.0215$$



from Tables corresponding Area = 0.1539

$\therefore$ Probability = 15.39%.

Jan 2017

7) Explain Normal Distribution. What are its Properties. Example 80 Explain Sampling Distribution.

A car Manufacturing factory has two plants. plant A manufactures 70% of the cars and plant B manufactures 30% of the cars. At plant A, 80% of the cars are rated as standard quality and at plant B, 90% of the cars are rated as standard quality. A car is picked up at random and is found to be of standard quality. What is the probability that it has come from plant A.

Sol Suppose the probability that the car picked up and the car come out from plant A.

$$\text{i.e. } P(A) = \frac{70}{100} = 0.7$$

Similarly the probability that the car has come from plant B

$$\text{i.e. } P(B) = \frac{30}{100} = 0.3$$

Let us suppose the probability that the car is of standard quality

The probability that the car has come from plant A = $\frac{P(S)}{P(A)}$

$$= P(S/A) = \frac{80}{100} = 0.8$$

Similarly the probability that the car has come from plant B

$$P(S/B) = \frac{90}{100} = 0.9$$

The probability that the car which is of standard quality is manufactured from plant A = $P(A/S)$

Now According to Baye's Theorem

$$\begin{aligned} P(A/S) &= \frac{P(A) \cdot P(S/A)}{P(A) \cdot P(S/A) + P(B) \cdot P(S/B)} \\ &= \frac{0.7 \times 0.8}{(0.7 \times 0.8) + (0.3 \times 0.9)} \\ &= \frac{0.56}{0.56 + 0.27} \Rightarrow \frac{0.56}{0.83} = 0.6747 \end{aligned}$$

- 2014
8) a) Define : i) probability ii) Marginal probability iii) condition probability.
iv) State and prove Addition and Multiplication laws of probability.
b) The chance that a female worker in a chemical factory will contract an occupational disease is 0.04 and the chance that for a male worker is 0.06. out of 800 workers in a factory, 150 are female. One worker is selected at random and the worker is found to have contracted the disease. What is the probability that the worker is a Male.

Sol

$$\text{No. of workers} = 800$$

$$\text{No. of Female workers} = 150$$

$$\text{Male workers} = 800 - 150 = 650$$

$$\text{Males} = \frac{650}{800} = 0.8125 \quad ; \quad \text{Females} = \frac{150}{800} = 0.1875$$

$$P(D/F) = 0.04, \quad P(D/M) = 0.06$$

The probability that when one worker is selected at random, the worker is a male workers

$$\begin{aligned} \Rightarrow P(M/D) &= \frac{P(M) \cdot P(D/M)}{P(M) \cdot P(D/M) + P(F) \cdot P(D/F)} \\ &= \frac{0.8125 \times 0.06}{(0.8125 \times 0.06) + (0.04 \times 0.1875)} \\ &= \frac{0.0488}{0.0563} = 0.8668 \\ &= 86.68\% \end{aligned}$$

1013
9) i) State Addition, Multiplication and Bayes Theorem. What are the properties of Normal Distribution.

ii) An Insurance Company Insured 1500 scooter drivers, 3000 car drivers and 6000 Truck Drivers. The Probability of an accident is 0.05, 0.02 and 0.10 respectively in case of scooter, car and Truck drivers. one of the Insured persons meets with an accident. What is the probability that
a) He is a scooter driver
b) He is a Truck driver

Sol
No. of Scooter Drivers Insured = 1500
No. of Car Drivers Insured = 3000
No. of Truck Drivers Insured = 6000

10,500

$$\text{Scooter drivers} \Rightarrow \frac{1500}{10,500} = 0.1428$$

$$\text{Car drivers} \Rightarrow \frac{3000}{10,500} = 0.2857$$

$$\text{Truck Drivers} \Rightarrow \frac{6000}{10,500} = 0.5714$$

Given that probability of a scooter accident = $P(A/S) = 0.05$

Given that probability of a car accident = $P(A/C) = 0.02$

Given that probability of a Truck Accident = $P(A/T) = 0.10$

The probability that when one of the insured persons meets with an accident

i) he is a scooter driver

According to Bayes Theorem

$$\begin{aligned} \Rightarrow P(S/A) &= \frac{P(S) \cdot P(A/S)}{P(S)P(A/S) + P(C)P(A/C) + P(T)P(A/T)} \\ &= \frac{(0.1428)(0.05)}{(0.1428 \times 0.05) + (0.02 \times 0.2857) + (0.10 \times 0.5714)} \\ &= \frac{0.0071}{0.0071 + 0.0057 + 0.0571} \Rightarrow \frac{0.0071}{0.0699} \end{aligned}$$

$$\boxed{P(S/A) = 0.1016}$$

ii) He is a Truck Driver

$$\begin{aligned}
 P(T/A) &= \frac{P(T) \cdot P(A/T)}{P(S) \cdot P(A/S) + P(C) \cdot P(A/C) + P(T) \cdot P(A/T)} \\
 &= \frac{0.10 \times 0.5714}{(0.1428 \times 0.05) + (0.02 \times 0.2857) + (0.10 \times 0.5714)} \\
 &= \frac{0.05714}{0.0071 + 0.0057 + 0.05714} \\
 &= \frac{0.05714}{0.0699}
 \end{aligned}$$

$$P(T/A) = 0.8175$$

iii) He is a car Driver

$$\begin{aligned}
 P(C/A) &= \frac{P(C) \cdot P(A/C)}{P(S) \cdot P(A/S) + P(C) \cdot P(A/C) + P(T) \cdot P(A/T)} \\
 &= \frac{0.02 \times 0.2857}{0.0071 + 0.0057 + 0.05714} \\
 &= \frac{0.0057}{0.0699}
 \end{aligned}$$

$$P(C/A) = 0.0817$$

^{sol²}
 10) A bag contain 4 white and 5 Black balls. Another Bag contains 4 white and 5 Black balls. From any one of these bags, a single draw of two balls is made. Find the probability that one of them would be white and another black ball

Sol

Bag I	Bag II
4W	4W
5B	5B
9	9

The probability that one of them would be white and other Black means, it may be noted that the first ball drawn may be white and the second ball drawn may be black and the opposite may also take place i.e first Black and second white.

$$\Rightarrow P(w_1) P(B_2) + P(B_1) P(w_2)$$

$$\Rightarrow \left(\frac{4}{9} \times \frac{5}{9}\right) + \left(\frac{5}{9} \times \frac{4}{9}\right)$$

$$\Rightarrow 0.2469 + 0.2469$$

$$= 0.4938$$

$$= 49.38\%$$

Monday
11/12/17

UNIT-V SAMPLING - TESTING OF HYPOTHESIS

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. The necessity of sampling is because it saves time and money.

Sampling Methods:-

Sampling methods are broadly classified as follows

1. Probability or Random Sampling.

2. Non-Probability or Non-Random Sampling.

1. Probability or Random Sampling:-

The Probability Sampling also called Random Sampling is based on the Theory of Probability that is it provides a known non-zero chance of selection for each population element.

For example - if a sample of 100 students is considered from a population of 10,000 students studying MBA, according to Random Sampling each student has $1/100^{\text{th}}$ chance of being selected.

2. Non Probability or Non Random Sampling:-

This sampling is not based on the Theory of Probability and hence this sampling does not provide a chance of selection for each population element.

Terminology:-

1. Null Hypothesis:- A definite statement about the population parameters for applying "Tests of Significance" is called Null Hypothesis which is denoted by " H_0 ".

For example:- If we want to decide whether one procedure or method is better than other procedure or method, we formulate

the Null hypothesis that there is no difference between the procedures or methods.

2. Alternative Hypothesis:- Any Hypothesis which is complementary to Null Hypothesis is called Alternative Hypothesis which is denoted by ' H_1 '.

For Example:- If we want to test the Null hypothesis that the average height of the soldiers in the Indian Army

$$H_0 = 165 \text{ cm} ; H_0: \mu = 165 \text{ cm}$$

$$H_1 \neq 165 \text{ cm} ; H_1: \mu \neq 165 \text{ cm}$$

$$\text{i.e., } \mu > 165 \text{ or } < 165$$

3. Test of Significance:- To test whether the null hypothesis setup should be "accepted or rejected", we use a test statistic which is based on appropriate probability distribution for this purpose

a) t -Distribution is used for small samples $n < 30$

b) z -Distribution is used for large samples $n \geq 30$

t -test $\rightarrow$ Small Sample Test.

z -test $\rightarrow$ Large Sample Test.

4. Type I Error:- If we reject the null hypothesis when it should be accepted, we say that Type I Error is committed. The probability of committing a Type I Error is denoted by ' α ' i.e.

$$P[\text{Reject } H_0 \text{ when it is true}] = \alpha.$$

5. Type II Error:- If we accept the null hypothesis when it should be rejected, we say that Type II Error is committed. The probability of committing a Type II Error is denoted by ' β '

$$\text{i.e., } P[\text{Accept } H_0 \text{ when it is false}] = \beta.$$

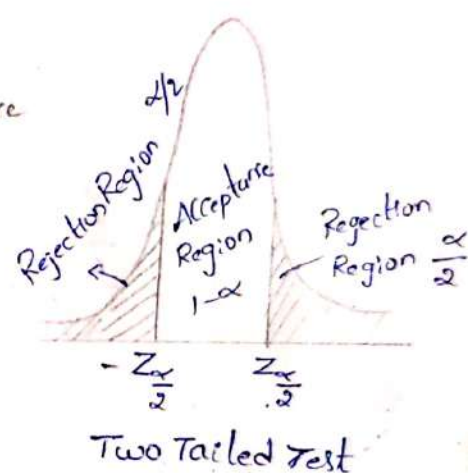
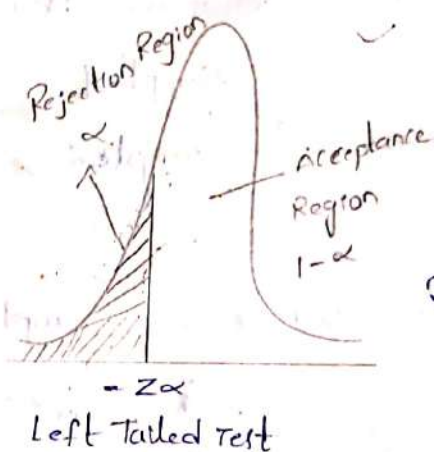
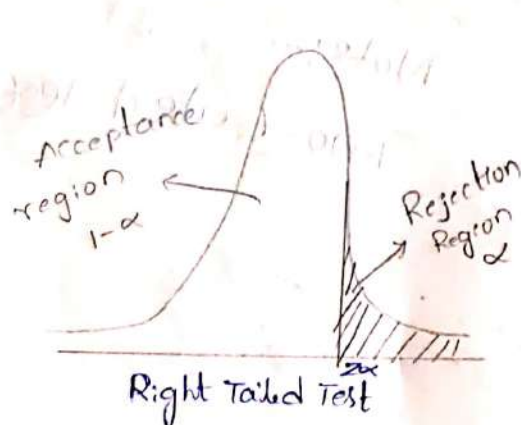
6. Level of Significance:- In testing of a given hypothesis, the maximum probability with which we would be willing to risk a Type I Error is called the level of Significance. In other words it is the

maximum probability of committing a Type I Error and is denoted by α !

7. One Tailed and Two Tailed Test:-

H₁ A Test of any statistical hypothesis where the alternative hypothesis is expressed by the symbol $\leq$ or $\geq$ is called One Tailed Test. Since the entire critical region lies in one Tail of the distribution of the Test Statistic, the critical region for alternative hypothesis containing the symbol ' $>$ ' has entirely Right Tail of the Distribution. The critical Region for an alternative hypothesis containing the symbol ' $<$ ' lies entirely in the Left Tail.

A Test of any statistical hypothesis where the alternative hypothesis is written with a symbol ' $\neq$ ' is called Two Tailed Test.



By comparing critical value with the calculated value, we accept or reject the null hypothesis at 5% or 1% level of significance.

Testing of Hypothesis - Procedure

Step ①: Setup the Null hypothesis (H_0)

Step ②: Setup the Alternative Hypothesis H_1 , which will enable us to decide whether to use One Tailed [Left/Right] Test or Two Tailed Test.

Step ③: choose the appropriate level of Significance ' α '

Step ④: compute the Test Statistic under the Null hypothesis using the following formula for Small samples and large samples.

problem 10
 $\leq$ "Left tailed test"
 $\geq$ "Right tailed test"
 $\neq$ "Two tailed test"

$$Z(\text{Single mean}) = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$Z(\text{Two means}) = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where $\bar{x}$ = Sample mean

μ = Population mean

σ = Standard Deviation of population.

n = Sample Size.

$\bar{x}_1$ = Mean of Sample 1.

$\bar{x}_2$ = Mean of Sample 2.

σ = Standard Deviation

n_1 = Sample 1.

n_2 = Sample 2.

Note:-
Two-tailed test

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

Where n = Sample Size.

$$\bar{x} = \frac{1}{n} \sum x_i$$

μ = Population mean.

S = Standard Deviation.

$$S = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

Where t -test is used for Small samples and z -test is used for large samples.

Step ⑤:- The calculated values of t and z are compared with the critical values (Table value) or Significant value at a given level of Significance as follows

Z -test

a) if $|z| \leq Z_{\alpha}$ i.e., the calculated value of z (modulus) is less than the critical value, then the Null hypothesis is accepted.

b) if $|Z| > Z_{\alpha}$, i.e. if the calculated value of Z is greater than the critical value, Null Hypothesis is Rejected

t-Test:-

a) if the calculated value $|t| \geq$ table value for $(n-1)$ degrees of freedom (df) at $\alpha\%$ level of significance, Reject H_0 .

b) if the calculated value $|t| <$ table value for $(n-1)$ degrees of freedom (df) at $\alpha\%$ level of significance, Accept H_0 .

Critical values of Z

Critical value Z_{α}	Level of Significance		
	1%	5%	10%
Two Tailed Test	$ Z_{\alpha/2} = 2.58$	$ Z_{\alpha/2} = \underline{1.96}$	$ Z_{\alpha/2} = 1.645$
Right Tailed Test	$Z_{\alpha} = 2.33$	$Z_{\alpha} = \underline{1.645}$	$Z_{\alpha} = 1.28$
Left Tailed Test	$-Z_{\alpha} = -\underline{2.33}$	$-Z_{\alpha} = -1.645$	$-Z_{\alpha} = -1.28$

$\begin{matrix} < & L \\ & e \\ > & R \end{matrix}$

Degrees of freedom (df):-

The Degrees of freedom is an important concept related to Hypothesis Testing. The Degrees of freedom in a distribution is the number of observations or values that are independent of each other that is the Number of values which we can choose freely.

Suppose there are six values in a sample with a mean of 18, In this case the degrees of freedom (df) or the number of values we can specify freely are $\boxed{6-1=5}$ i.e., to generalise, if the sample size is ' n ', the degrees of freedom is $\boxed{n-1}$.

Tuesday 12/12/17

1) A random Sample of 400 flower stems has an average length of 10cm. can this be regarded as a sample drawn from a large population with a mean of 10.2cm and standard deviation of 2.25cm

Sol

Given that $n = 400$;

mean of the Sample $\bar{x} = 10\text{cm}$

Mean of the population $\mu = 10.2\text{cm}$

Standard Deviation $\sigma = 2.25\text{cm}$

H_0 : The Sample has been drawn from the Normal population with mean $\mu = 10.2\text{cm}$.

H_1 : The Sample has not been drawn from the Normal population. $\mu \neq 10.2\text{cm}$.

Hence it is decided to conduct two tail test.

$H_0: \mu = 10.2$

$H_1: \mu \neq 10.2$

Computation of Test statistic

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{10 - 10.2}{2.25 / \sqrt{400}} =$$

$=$

$$= \frac{-0.2}{\frac{2.25}{20}}$$

$$= \frac{-0.2}{0.1125} = -1.7778$$

$$|Z| = 1.77$$

Assuming Level of Significance as 5%, we have critical value of Z

for Two-Tail Test $\Rightarrow Z_{cr} = 1.96$ ✓

$$Z_{cal}(1.77) < Z_{cr}(1.96)$$

So, Accept null Hypotheses.

Hence the Sample can be assumed to be drawn from the Normal population with mean = 10.2cm and standard deviation 2.25cm.

2) The mean life-time of a Sample of 100 fluorescent light tubes produced by a company is computed to be 1570 hours with a standard deviation of 120 hrs. The company claims that the Average life of the tubes produced by the company is 1600 hrs. using the level of Significance of

Two tailed test
दो ओर का परीक्षण
जोड़/घटाव

0.05, is the claim acceptable

Sol

Given that $n = 100$

Mean of the sample $\bar{x} = 1570$ hrs

Mean of the population $\mu = 1600$ hrs

Standard Deviation $\sigma = 120$ hrs.

H_0 : The Sample has been drawn from the company with $\mu = 1600$ hrs.

H_1 : The Sample has not been drawn from the company $\mu \neq 1600$ hrs.

$H_0: \mu = 1600$

$H_1: \mu \neq 1600$

Hence to conduct large sample Test $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{1570 - 1600}{120/\sqrt{100}}$$

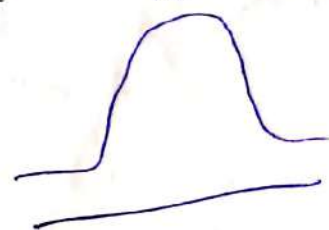
$$= \frac{-30}{\frac{120}{10}} = \frac{-30}{12} = -2.5$$

$$|Z| = 2.5$$

At 0.05 Level of Significance $Z_{cr} = 1.96$ ✓

We observe $Z_{cal}(2.5) > Z_{cr}(1.96)$.

So H_0 Rejected.



3) Suppose that 100 Tyres of a certain brand lasted on the average 21,431 miles with a Standard Deviation of 1295 miles using $\alpha = 0.05$, Test the Null hypothesis $\mu = 22000$ miles against alternative hypothesis $\mu < 22000$ miles

Sol

Given that $H_0: \mu = 22000$

$H_1: \mu < 22000$

$$\bar{x} = 21431$$

$$\sigma = 1295$$

Hence Test statistic $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{21431 - 22000}{1295/\sqrt{100}} = \frac{-569}{129.5} = -4.39$$

$$|Z| > 4.39 \quad Z_{cal} = -4.39$$

$$Z_{cr} = -1.645$$

We observe $Z_{cal}(-4.39) < Z_{cr}(-1.645)$, So H_0 Accepted.

und. 3000
left tailed test
This test left
tailed test left
tailed test
modulus of 4.39

4) The means of two large samples of sizes 1000 and 2000 are 67.5 and 68.0 respectively. Test the equality of means of the two populations each with a standard deviation of 2.5

Sol $n_1 = 1000$; $n_2 = 2000$
 $\bar{x}_1 = 67.5$; $\bar{x}_2 = 68.0$; $\sigma_1 = \sigma_2 = 2.5 = \sigma$

H₀: $\mu_1 = \mu_2$

H₁: $\mu_1 \neq \mu_2$

Hence decided Two Tailed Test

Test Statistic $\Rightarrow Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$= \frac{67.5 - 68}{2.5 \sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= \frac{-0.5}{2.5 \sqrt{0.0010 + 0.0005}}$$

$$= \frac{-0.5}{2.5(0.0387)} = \frac{-0.5}{0.0968}$$

$|Z| = 5.16$

$Z = -5.16$

At 5% level of significance $Z_{\text{cri}} = 1.96$

$Z_{\text{cal}} (5.16) > Z_{\text{cri}} (1.96)$

Hence Rejected H₀

Left tailed test

t-Test Small Sample Test

1) In a 12 Test runs over a marked course, a newly designed motor boat averages 33.6 Secs with a standard deviation of 2.3 Secs. Assuming that it is reasonable to treat the data as a random sample from a Normal population, Test the Null hypothesis $\mu = 35$ against Alternative hypothesis $\mu < 35$ at the level of significance $\alpha = 0.05$

Sol Given that $H_0: \mu = 35 \text{ sec}$; $\bar{x} = 33.6$; $n = 12$
 $H_1: \mu < 35 \text{ sec}$; $\sigma = 2.3$; $\mu = 35$

Test statistics

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{33.6 - 35}{2.3/\sqrt{12}} \Rightarrow \frac{-1.4}{2.3/3.4641} = \frac{-1.4}{0.6670} = -2.12$$

$$t_{cal} = -2.12$$

The critical Region is in the left tail

from the tables

$$t_{0.05, 11} = -1.796$$

$$df = n - 1 \Rightarrow 12 - 1 = 11$$

We observe

$$t_{cal} (-2.12) < t_{cr} \text{ or } t_{0.05, 11} (-1.796)$$

Hence Accepted.

2) A machine which produces Mica insulating washers for use in electric devices is set to turn out washers having a thickness of 10mm. A sample of 10 washers has an Average thickness of 9.52 mm with a standard deviation of 0.60 mm. Test whether the sample is drawn from the given population.

Two tailed test

Sol Given that $n = 10$; $\sigma = 0.60$
 $\bar{x} = 9.52$; $\mu = 10$

$$H_0: \mu = 10$$

$$H_1: \mu \neq 10$$

Test statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \Rightarrow \frac{9.52 - 10}{0.60/\sqrt{10}} \Rightarrow \frac{-0.48}{0.60/3.1623}$$

$$t_{cal} = -2.53$$

$$= \frac{-0.48}{0.1897}$$

of test is < 2.12 so left tailed test is accepted

$$1 - \alpha = 1 - 0.05 = 0.95$$

Two tailed test is accepted

Because it is two-tailed Test

$$1-\alpha = 2.53$$

Given $n=10$

$$df = n-1 = 10-1 = 9$$

At 5% (0.05) Level of Significance

$$t_{0.05, 9} = 1.833$$

We observe $t_{cal}(2.53) > t_{0.05, 9}(1.833)$

Hence Rejected.

So, the Sample is not drawn from the population.

3) A random sample of 10 boys had the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Does this data support the assumption of a population mean IQ of 100.

Sol

$$H_0: \mu = 100$$

$$H_1: \mu \neq 100$$

Test Statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum x_i \\ &= \frac{972}{10} \\ &= 97.2 \end{aligned}$$

$$S = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

$$S = \sqrt{\frac{1}{10-1} \times 1833.60}$$

$$S = \sqrt{\frac{1}{9} \times 1833.60}$$

$$S = \sqrt{203.73}$$

$$S = 14.27$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84

$$\sum x = 972 \quad \sum (x - \bar{x})^2 = 1833.60$$

Label value
with μ_0 does
Rejected
with μ_0 does
Accepted

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{97.2 - 100}{14.27/\sqrt{10}} = \frac{-2.8}{14.27/3.1623} = \frac{-2.8}{4.5125} = -0.62$$

Because it is two-tailed test $|t| = 0.62$

$$|t| = |-0.62|$$

$$t_{0.05, 9} = 1.833$$

$$t = 0.62$$

We observe

$$t_{cal}(0.62) < t_{cr}(1.833)$$

Hence Accepted.

*) Test of Significance of difference b/w two Means (small sample) :-

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

tending
↓
degrees of freedom.

Where $n_1 = 1^{st}$ sample size ; $\bar{x}_1 =$ mean of the first sample
 $n_2 = 2^{nd}$ sample size ; $\bar{x}_2 =$ Mean of 2nd sample

$$S^2 = \frac{1}{n_1+n_2-2} \left[(n_1-1)S_1^2 + (n_2-1)S_2^2 \right]$$

$$S_1^2 = \frac{1}{(n_1-1)} \sum (x_i - \bar{x}_1)^2$$

$$S_2^2 = \frac{1}{(n_2-1)} \sum (x_i - \bar{x}_2)^2$$

4) To Examine the hypothesis that Husbands are more intelligent than wives, an Investigator took a sample of 10 couples and conducted a test which measures the IQ as follows

Husbands: 117 105 97 105 123 109 86 78 107 107

Wives: 106 98 87 104 116 95 90 69 108 85

Test the hypothesis with a reasonable test at 0.05 level of significance.

Sol Given $n_1 = 10$; $n_2 = 10$

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

$\bar{x}_1$	x_i	$(x_i - \bar{x}_1)$	$(x_i - \bar{x}_1)^2$	$(x_j - \bar{x}_2)$	$(x_j - \bar{x}_2)^2$
117	106	13.6	184.96	10.2	104.04
105	98	2.56 1.6	2.56	2.2	4.84
97	87	-6.4	40.96	-8.8	77.44
105	104	1.6	2.56	8.2	67.24
123	116	19.6	384.16	20.2	408.04
109	95	5.6	31.36	-0.8	0.64
86	90	-17.4	302.76	-5.8	33.64
78	69	-25.4	645.16	-26.8	718.24
107	108	3.6	12.96	12.2	148.84
107	85	3.6	12.96	-10.8	116.64
<u>1034</u>	<u>958</u>		<u>1620.4</u>		<u>1679.6</u>

$$\bar{x}_1 = \frac{\sum x_i}{n} = \frac{1034}{10} = 103.4$$

$$\bar{x}_2 = \frac{\sum x_j}{n} = \frac{958}{10} = 95.8$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum (x_i - \bar{x}_1)^2$$

$$= \frac{1}{10 - 1} \times 1620.4 = \frac{1}{9} \times 1620.4$$

$$= 180.04$$

$$\Rightarrow S_1 = \sqrt{180.04}$$

$$= 13.42$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum (x_j - \bar{x}_2)^2$$

$$= \frac{1}{10 - 1} \times 1679.6 \Rightarrow \frac{1}{9} \times 1679.6 = 186.62$$

$$\Rightarrow S_2 = \sqrt{186.62} = 13.66$$

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2 \right]$$

$$S^2 = \frac{1}{(10+10-2)} \left[(10-1)(180.04) + (10-1)(186.62) \right]$$

$$S^2 = \frac{1}{18} \left[(9 \times 180.04) + (9 \times 186.62) \right]$$

$$S^2 = \frac{1}{18} [1620.36 + 1679.58]$$

$$S^2 = \frac{3299.94}{18}$$

$$S^2 = 183.33$$

$$S = \sqrt{183.33}$$

$$S = 13.54$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow \frac{103.4 - 95.8}{13.54 \sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$\Rightarrow \frac{7.6}{13.54 \sqrt{0.1 + 0.1}} \Rightarrow \frac{7.6}{13.54 \times 0.4472}$$

$$\Rightarrow \frac{7.6}{6.0553} \Rightarrow t = 1.25$$

$$t_{cal} = 1.25$$

critical value

$$t_{n_1+n_2-2}$$

$$\Rightarrow t_{10+10-2} = t_{18}$$

$$t_{0.05, 18} = 1.734$$

We observe $t_{cal} (1.25) < t_{0.05, 18} (1.734)$

So, Null Hypothesis Accepted

5) The Time taken by workers in performing a job by Method I and Method II are given below

Method I 20 16 26 27 23 22

Method II 27 33 42 35 32 34 38

Do the data show that the ^{means} variances of Time Distribution from population from which these samples are drawn do not differ significantly.

^{In 2016} 6) Two kinds of Manures were applied to ~~16~~ ^{1000 acres} sixteen one hectare plots. Other conditions remaining the same, the yields in quintals are given below

Manure I 18 20 36 50 49 36 34 49 41

Manure II 29 28 26 35 30 44 46

Is there any significant difference between the mean yield.

Sol

*** Chi-Square Test (χ^2) (774 page Table)

* chi-square criteria are based on observed and expected cell frequencies. This test plays an important role in many problems dealing with the analysis of count data in Business and Industry.
(No. of +, -, x, ÷, +)

→ The rxc tables in chi-square test is called contingency table.

→ The rxc tables, Association is found between two attributes A & B under consideration. If the hypothesis is rejected, then the two attributes are considered to be associated, otherwise they are considered Independent.

For Example :- Whether a given medicine 'A' is effective in curing a disease 'B', Whether a given fertiliser 'A' is effective in increasing the crop yield 'B' (or) whether the advertisement A has impact on the sales 'B'.

→ In chi-square test, the rows and columns table which is called contingency table is as follows.

	A_1	A_2		A_c	
B_1	$A_1 B_1$	$A_2 B_1$		$A_c B_1$	B_1
B_2	$A_1 B_2$	$A_2 B_2$		$A_c B_2$	B_2
B_r	$A_1 B_r$	$A_2 B_r$		$A_c B_r$	B_r
	A_1	A_2		A_c	(N)

* In chi-square test, the frequencies given in the cells are called Observed frequencies ('O') and the frequencies to be calculated are called Expected frequencies.

* The formula used to calculate the Expected frequencies

$$e(A_1, B_1) = \frac{(A_1 \cdot B_1)}{N}$$

$$e(A_2, B_1) = \frac{(A_2 \cdot B_1)}{N}$$

$$e(A_c, B_r) = \frac{(A_c \cdot B_r)}{N}$$

So, finally

$$\chi^2_{cal} = \sum_{i=1}^r \sum_{j=1}^c \frac{[O_{ij} - e_{ij}]^2}{e_{ij}}$$

$$\chi^2_{crit} \sim (r-1)(c-1) df$$

Model - I

1) In an experiment on Immunization of cattle from tuberculosis the following particulars are available.

	<u>Affected</u>	<u>Not-affected</u>
Inoculated	12	26
Non-Inoculated	16	6

Analyse and Discuss the effect of vaccine in controlling the tuberculosis.

Sol

Let H_0 : The variables are independent (vaccine & Tuberculosis)

H_1 : The variables are dependent.

12	26	38
16	6	22
28	32	60

Observed cell frequency

$$O_{11} = 12$$

$$O_{12} = 26$$

$$O_{21} = 16$$

$$O_{22} = \frac{6}{60}$$

Expected frequencies

$$e_{11} = \frac{38 \times 28}{60} = 17.73$$

$$e_{12} = \frac{38 \times 32}{60} = 20.27$$

$$e_{21} = \frac{22 \times 28}{60} = 10.27$$

$$e_{22} = \frac{22 \times 32}{60} = 11.73$$

$$\begin{aligned} \text{Now } \chi^2_{\text{cal}} &= \frac{(12 - 17.73)^2}{17.73} + \frac{(26 - 20.27)^2}{20.27} + \frac{(16 - 10.27)^2}{10.27} + \frac{(6 - 11.73)^2}{11.73} \\ &= \frac{(-5.73)^2}{17.73} + \frac{(5.73)^2}{20.27} + \frac{(5.73)^2}{10.27} + \frac{(-5.73)^2}{11.73} \\ &= \frac{32.83}{17.73} + \frac{32.83}{20.27} + \frac{32.83}{10.27} + \frac{32.83}{11.73} \\ &= 1.85 + 1.62 + 3.19 + 2.79 \end{aligned}$$

$$\boxed{\chi^2_{\text{cal}} = 9.43}$$

$$\chi^2_{0.05, (r-1)(c-1)} = \chi^2_{(0.05) (2-1) (2-1)}$$

$$= \chi^2_{0.05, 1} = 3.841$$

$$\boxed{\chi^2_{\text{cri}} = 3.841}$$

condition

$$\chi^2_{\text{cal}} < \chi^2_{\text{cri}} \text{ (Accept)}$$

$$\chi^2_{\text{cal}} > \chi^2_{\text{cri}} \text{ (Reject)}$$

We observe

$$\chi^2_{\text{cal}} (9.43) > \chi^2_{\text{cri}} (3.841)$$

So, H_0 Rejected

i.e. H_1 is accepted

H_0 is now rejected

There exist an Association between vaccine & Tuberculosis.

2) In an anti-malarial campaign in certain area, Quinine was given for a population in that area and the no. of fever cases recorded are given below

	fever	No-fever
Quinine :	20	792
No quinine :	220	2216

Discuss the usefulness of Quinine.

3) The following data represents a person's ability in mathematics and his interest in statistics.

		Ability in mathematics		
		Low	Avg	High
Interest in Statistics:	Low	63	42	15
	Avg	58	61	31
	High	14	47	29

Use 0.01 level of significance and test for Independence b/w person's ability in maths and in statistics.

Sol Let H_0 = They are independent

H_1 = They are dependent

63	42	15	120
58	61	31	150
14	47	29	90
135	150	75	360

Cell frequency

Expected frequency

$$O_{11} = 63$$

$$e_{11} = \frac{120 \times 135}{360} = 45$$

$$O_{12} = 42$$

$$e_{12} = \frac{120 \times 150}{360} = 50$$

$$O_{13} = 15$$

$$e_{13} = \frac{120 \times 75}{360} = 25$$

$$O_{21} = 58$$

$$e_{21} = \frac{150 \times 135}{360} = 56.25$$

$$O_{22} = 61$$

$$e_{22} = \frac{150 \times 150}{360} = 62.5$$

$$O_{23} = 31$$

$$O_{31} = 14$$

$$O_{32} = 47$$

$$O_{33} = 29$$

$$e_{23} = \frac{150 \times 75}{360} = 31.25$$

$$e_{31} = \frac{90 \times 135}{360} = 33.75$$

$$e_{32} = \frac{90 \times 150}{360} = 37.5$$

$$e_{33} = \frac{90 \times 75}{360} = 18.75$$

$$\chi^2_{cal} = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(63-45)^2}{45} + \frac{(42-50)^2}{50} + \frac{(15-25)^2}{25} + \frac{(58-56.25)^2}{56.25} + \frac{(61-62.5)^2}{62.5} + \frac{(31-31.25)^2}{31.25} \\ + \frac{(14-33.75)^2}{33.75} + \frac{(47-37.5)^2}{37.5} + \frac{(29-18.75)^2}{18.75}$$

$$= \frac{324}{45} + \frac{64}{50} + \frac{100}{25} + \frac{4}{56.25} + \frac{2.25}{62.5} + \frac{0.0625}{31.25} + \frac{390.1}{33.75} + \frac{90.25}{37.5} + \frac{105.1}{18.75}$$

$$= 7.2 + 1.28 + 4 + 0.036 + 0.002 + 11.56 + 2.41 + 5.61$$

$$\boxed{\chi^2_{cal} = 32.10}$$

$$\chi^2_{crit(0.01)(3-1)(3-1)} = \chi^2_{crit(0.01)(3-1)(3-1)} \Rightarrow \chi^2_{crit(0.01, 4)} = 13.277$$

$$* \chi^2_{cal} (32.10) > \chi^2_{crit} (13.277) \rightarrow \text{Rejected } H_0 \text{ \& accept } H_1.$$

So, the ability of Maths and Interest in statistic of person is dependent.

Model-II

chi-square Test of Goodness of fit

* chi-square test is used to check the significance of the discrepancy b/w Theory and Experiment. It is given by

Karl Pearson (P) in 1900 and is popularly known as "chi-square Test of goodness of fit".

If O_i ($i = 1, 2, 3, 4, \dots, n$) is a set of observed (or) Experimental frequencies and e_i ($i = 1, 2, 3, \dots, n$) is a set of expected frequencies (or) Theoretical frequencies

chi-square test is given by

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{(n-1)} \text{ df}$$

1) find the value of chi-square for the following data and test the goodness of fit.

Observed frequency : 14, 15, 18, 20, 15, 10

Expected frequency : 17, 10, 15, 25, 10, 15

Sol H_0 : The given distribution fit is good

H_1 : The given distribution fitness of good does not exist

$$\chi^2_{\text{cal}} = \sum_{i=1}^n \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{(n-1)} \text{ df}$$

$$= \frac{(14-17)^2}{17} + \frac{(15-10)^2}{10} + \frac{(18-15)^2}{15} + \frac{(20-25)^2}{25} + \frac{(15-10)^2}{10} + \frac{(10-15)^2}{15}$$

$$= \frac{9}{17} + \frac{25}{10} + \frac{9}{15} + \frac{25}{25} + \frac{25}{10} + \frac{25}{15}$$

$$= 0.53 + 2.5 + 0.6 + 1 + 2.5 + 1.67$$

$$\boxed{\chi^2_{\text{cal}} = 8.8}$$

$$\text{df} = n-1 = 6-1 = 5$$

$$\chi^2_{\text{cri}(0.05, 5)} = 11.070$$

$\chi^2_{\text{cal}} (8.8) < \chi^2_{\text{cri}} (11.070)$. It is accepted H_0 so given distribution fit is good.

2) A personnel manager is interested in trying to determine whether absenteeism is greater on one day of the week than on the other day of the week. He has the following record for the past year.

Days of Week : mon Tues Wed Thurs Fri

No. of Absentees : 66 57 54 48 75

Test whether Absentism is uni-formly distributed over the week. By using test of goodness of fit.

Sol H_0 : They are uni-formly Distributed (fit is good)

H_1 : They are not uni-formly Distributed (fit is not good).

* Expected frequency = $\frac{66+57+54+48+75}{5} = \frac{300}{5} = 60$

uni-formly Absentees

$$\chi^2_{cal} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{df}$$

$$= \frac{(66-60)^2}{60} + \frac{(57-60)^2}{60} + \frac{(54-60)^2}{60} + \frac{(48-60)^2}{60} + \frac{(75-60)^2}{60}$$

$$= \frac{36}{60} + \frac{9}{60} + \frac{36}{60} + \frac{144}{60} + \frac{225}{60}$$

$$= \frac{450}{60} = 7.5$$

$$\chi^2_{cal} = 7.5$$

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
66	60	6	36	0.6
57	60	-3	9	0.15
54	60	-6	36	0.6
48	60	-12	144	2.4
75	60	15	225	3.75
				7.5

$$\chi^2_{crit} = \chi^2_{0.05(5-1)} = 9.488$$

$$\chi^2_{cal}(7.5) < \chi^2_{crit}(9.488)$$

H_0 is Accepted

Hence there is uni-formly Distribution of absence/Week fit of good exist.

3) The following figures show the Distribution of digits in numbers chosen at random from a telephone directory

Digit : 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

frequency : 1026, 1107, 997, 966, 1075, 933, 1107, 972, 964, 853

Test whether the digits may be taken to occur equally frequently in the Directory.

Sol H_0 : They are uni-formly (equally) -frequent

H_1 : They are not uni-formly -frequent.

Expected frequency = $\frac{\sum f}{n} = \frac{10000}{10} = 1000$

$$\chi^2_{cal} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(0.05) df}$$

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2$
1026	1000	26	676	0.676
1107	1000	107	11449	11.449
997	1000	-3	9	0.009
966	1000	-34	1156	1.156
1075	1000	75	5625	5.625
933	1000	-67	4489	4.489
1107	1000	107	11449	11.449
972	1000	-28	784	0.784
964	1000	-36	1296	1.296
853	1000	-147	21609	21.609
			<u>58542</u>	<u>58.542</u>

$$\chi^2_{cal} = \frac{58542}{1000} = 58.542$$

$$\chi^2_{cri} = \chi^2_{0.05(10-1)} \Rightarrow \chi^2_{0.05, 9} = 16.919$$

$$\chi^2_{cal} (58.542) > \chi^2_{cri} (16.919)$$

Wednesday
13/12/17

4) The following is the distribution of the hourly number of Trucks arriving at a companies warehouse

Trucks arrivals:	0	1	2	3	4	5	6	7	8
frequency	52	151	130	102	45	12	5	1	2

fit a poisson distribution and Test the Goodness of fit at 0.05 level of Significance.

Sol

 H_0 : The fit is good. H_1 : The fit is not good.

x_i	f_i	fx_i
0	52	0
1	151	151
2	130	260
3	102	306
4	45	180
5	12	60
6	5	30
7	1	7
8	2	16
	<u>500</u>	<u>1010</u>

$$\text{Mean} = \frac{\sum fx_i}{\sum f_i}$$

$$= \frac{1010}{500}$$

$$= 2.02$$

In poisson distribution

$$\lambda = \text{mean}$$

$$\lambda = 2.02$$

	$P(x) = N \cdot \frac{\lambda^x e^{-\lambda}}{x!}$	Expected frequency
0	$\frac{500 \cdot (2.02)^0 \cdot e^{-2.02}}{0!} = 500 \times 0.1327$	66
1	$\frac{500 \cdot (2.02)^1 \cdot e^{-2.02}}{1!} = 500 \times 2.02 \times 0.1327$	134
2	$\frac{500 \cdot (2.02)^2 \cdot e^{-2.02}}{2!} = \frac{500 \times 4.0804 \times 0.1327}{2!}$	135
3	$\frac{500 \cdot (2.02)^3 \cdot e^{-2.02}}{3!} = \frac{500 \times 8.2424 \times 0.1327}{6}$	91
4	$\frac{500 \cdot (2.02)^4 \cdot e^{-2.02}}{4!} = \frac{500 \times 16.6497 \times 0.1327}{24}$	46
5	$\frac{500 \cdot (2.02)^5 \cdot e^{-2.02}}{5!} = \frac{500 \times 33.6323 \times 0.1327}{120}$	19
6	$\frac{500 \cdot (2.02)^6 \cdot e^{-2.02}}{6!} = \frac{500 \times 67.9323 \times 0.1327}{720}$	6
7	$\frac{500 \cdot (2.02)^7 \cdot e^{-2.02}}{7!} = \frac{500 \times 137.23 \times 0.1327}{5040}$	2
8	$\frac{500 \cdot (2.02)^8 \cdot e^{-2.02}}{8!} = \frac{500 \times 277.211 \times 0.1327}{40320}$	1

O_i	e_i	$O_i - e_i$	$(O_i - e_i)^2$	$(O_i - e_i)^2 / e_i$
50	66	-14	196	2.97
151	134	17	289	2.16
130	135	-5	25	0.19
109	91	11	121	1.33
45	46	-1	1	0.02
12	19	-7	49	2.57
5	6	-1	1	0.17
1	2	-1	1	0.50
2	1	1	1	1
				10.90

$$\chi^2_{cal} = 10.90$$

$$\chi^2_{crit} \Rightarrow \chi^2_{0.05, 8} = 15.507$$

(9-1)

$$\chi^2_{cal} < \chi^2_{crit}$$

Accept H_0 .

The given distribution fit is good.

ANNOVA (Analysis of Variance) :-

* The Analysis of Variance was developed by Fisher. ANOVA helps us to Test the Significance of the differences of variances among more than two sample means. Using ANOVA, we can make inferences about whether the samples are drawn from the population having same mean. ANOVA is used in such situations as comparing the mileage given by (5) five different brands of fuel, comparing the first year earnings of Graduates of half-a-dozen business schools.

In ANOVA comparison will be made between means of more than two samples. ANOVA is the classification and cross classification of statistical data with a view of testing whether the mean of specific classification differ significantly or they are homogeneous.

ANOVA TABLE

The Technique of ANOVA includes a Table showing the Source of Variance, the Sum of Squares, the degrees of freedom, mean square (variance) and the formula for F-ratio

$$F_{\text{statistic}} = \frac{\text{Variance b/w the samples}}{\text{Variance within the samples}}$$

ANOVA is classified into

1. one way classification — one way ANOVA.
2. Two way classification — Two way ANOVA.
(Manifold ANOVA)

one way ANOVA:-

The simplest form of ANOVA is a one way model which is used with simple random samples in order to compare the effect of single independent variable on the dependent variable. The following is the specimen (model) of ANOVA Table.

Source of Variation	Sum of Squares	df	Mean Square	F-ratio
b/w samples	SSC	K-1	$MSC = \frac{SSC}{K-1}$ columns ←	$\frac{MSC}{MSE}$
Within Sample	$SSE (SST - SSC)$	N-K	$MSE = \frac{SSE}{N-K}$	
Total	SST	N-1 ↓ rows x columns		

1) Below are given the yield (Kgs per acre) for five trial plots of varieties of treatment

Plot NO	Treatment			
	1	2	3	4
1	42	48	68	80
2	50	66	52	94
3	62	68	76	78
4	34	78	64	82
5	52	70	70	66

check the hypothesis whether the treatments have the same effect.

Sol

plot	Treatment			
	X ₁	X ₂	X ₃	X ₄
1	42	48	68	80
2	50	66	52	94
3	62	68	76	78
4	34	78	64	82
5	52	70	70	66
Total	240	330	330	400

Let $T = \text{Sum of all observations}$

$$\Rightarrow 42 + 50 + 62 + 34 + 52 + 48 + 66 + 68 + \dots + 66 = 1300$$

$$\text{Now } \frac{T^2}{N} = \frac{(1300)^2}{5 \times 4}$$

$$N = r \times c$$

$$= 84500 \text{ (correction factor)}$$

$SST = \text{Sum of squares of all observations} - \text{correction factor}$

$$= [42^2 + 50^2 + 62^2 + 34^2 + 52^2 + \dots + 66^2] - 84500$$

$$= 88736 - 84500$$

$$SST = 4236$$

$$SSC = \frac{(\sum x_1)^2}{N_1} + \frac{(\sum x_2)^2}{N_2} + \frac{(\sum x_3)^2}{N} + \dots + \frac{(\sum x_n)^2}{N_n} - \frac{T^2}{N}$$

$$= \frac{(240)^2}{5} + \frac{(330)^2}{5} + \frac{(330)^2}{5} + \frac{(400)^2}{5} - 84500$$

$$= \frac{57600}{5} + \frac{108900}{5} + \frac{108900}{5} + \frac{160000}{5}$$

$$= 11520 + 21780 + 21780 + 32000 - 84500$$

$$= 87080 - 84500$$

$$SSC = 2580$$

$$SSE = SST - SSC$$

$$= 4236 - 2580 = 1656$$

$$MSC = \frac{SSC}{\underset{\text{columns}}{K-1}} = \frac{2580}{4-1} = 860$$

$$MSE = \frac{SSE}{N-K} = \frac{1656}{20-4} = 103.5$$

Source of Variation	Sum of Squares	df	Mean Square	F _{cal}
b/w Samples	SSC = 2580	K-1 = 3	MSC = 860	$\frac{MSC}{MSE} = \frac{860}{103.5} = 8.3$
Within Sample	SSE = 1656	N-K = 16 (20-4)	MSE = 103.5	
Total	SST	N-1		

$$F_{cal} = 8.3$$

$$df \Rightarrow (K-1), (N-K)$$

$$\Rightarrow (4-1), (20-4) \Rightarrow (3, 16)$$

$$\Rightarrow F_{table} = 2.24$$

$$F_{cal} > F_{table}$$

H₀ Rejected.

Thursday
14/12/17

Two-way ANOVA

In Two-way ANOVA, the data is classified according to two criteria.

Procedure for Two-way ANOVA:-

step ①: Assuming that the means of all columns is equal that is the effect of all factors in one kind of Treatment are equal i.e.,

$$\alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_c \quad (c - \text{columns})$$

step ②:- Assuming the means of all rows are equal that is the effect of all factors in the second Treatment are equal i.e.,

$$\beta_1 = \beta_2 = \beta_3 = \dots = \beta_r \quad (r - \text{rows})$$

step ③: compute $T = \text{Sum of all values}$

step ④: compute $SST = \text{Sum of all squares of observations} - \frac{T^2}{N}$

step ⑤: compute

$$SSC = \left(\frac{(\sum x_1)^2}{N_1} + \frac{(\sum x_2)^2}{N_2} + \dots + \frac{(\sum x_n)^2}{N_n} - \frac{T^2}{N} \right) -$$

where $\sum x_1, \sum x_2, \dots, \sum x_n$ are column Totals.

step ⑥: compute $SSR = \left(\frac{(\sum x_{j1})^2}{N_1} + \frac{(\sum x_{j2})^2}{N_2} + \dots + \frac{(\sum x_{jn})^2}{N_n} - \frac{T^2}{N} \right) -$

where $\sum x_{j1}, \sum x_{j2}, \dots, \sum x_{jn}$ are Row totals.

step ⑦:- $SSE = SST - SSC - SSR$

step ⑧:- $MSC = \frac{SSC}{c-1}$, $MSR = \frac{SSR}{r-1}$, $MSE = \frac{SSE}{(r-1)(c-1)}$

step ⑨:- $F_c = \frac{MSC}{MSE}$, $F_r = \frac{MSR}{MSE}$

step 10:- Degrees of freedom for $F_c = (c-1), (c-1)(r-1)$
 $F_r = (r-1), (c-1)(r-1)$

F_c = column wise comparison

F_r = Row wise comparison

Source of Variation	Sum of Squares	df	Mean Square	F-ratio
Columns	SSC	$c-1$	MSC	$F_c = \frac{MSC}{MSE}$
Rows	SSR	$r-1$	MSR	
Error	SSE	$(c-1)(r-1)$	MSE	$F_r = \frac{MSR}{MSE}$
Total	SST	$N-1$		

1) Three varieties of crops A, B, C are tested with four replications as given below. The yields are shown in the following Table.

Replications				
Variety	1	2	3	4
A	6	4	8	6
B	7	6	6	9
C	8	5	10	9

Test whether there is a difference between Replications and also the crop varieties differ significantly.

Sol

Replications					
Variety	1	2	3	4	
A	6	4	8	6	24
B	7	6	6	9	28
C	8	5	10	9	32
Total	21	15	24	24	84

$C-H_0$: There is a significant difference in replications

$R-H_0$: There is a significant difference in variety of crops.

Let $T =$ Sum of all observations
 $= 84$

$N =$ Number of observations $= 12$

$$\frac{T^2}{N} = \frac{84^2}{12} = \frac{7056}{12} = 588$$

$$SST = \text{Sum of squares of all observations} - \text{Correction factor } \frac{T^2}{N}$$

$$= (6^2 + 7^2 + 8^2 + 4^2 + 6^2 + 5^2 + 8^2 + 6^2 + 10^2 + 9^2 + 6^2 + 9^2) - 588$$

$$= 624 - 588$$

$$\boxed{SST = 36}$$

$$SSC = \frac{(\sum x_1)^2}{N_1} + \frac{(\sum x_2)^2}{N_2} + \frac{(\sum x_3)^2}{N_3} - \frac{T^2}{N}$$

$$= \frac{21^2}{3} + \frac{15^2}{3} + \frac{24^2}{3} + \frac{24^2}{3} - 588$$

$$= \frac{606}{3} - 588$$

$$\boxed{SSC = 18}$$

$$SSR = \frac{24^2}{4} + \frac{28^2}{4} + \frac{32^2}{4} - \frac{T^2}{N}$$

$$= 596 - 588$$

$$= 8 \Rightarrow SSR = 8$$

$$SSE = SST - SSC - SSR$$

$$= 36 - 18 - 8$$

$$= 10$$

$$MSC = \frac{SSC}{c-1} = \frac{18}{4-1} = \frac{18}{3} = 6$$

$$MSR = \frac{SSR}{r-1} = \frac{8}{3-1} = \frac{8}{2} = 4$$

$$MSE = \frac{SSE}{(c-1)(r-1)}$$

$$= \frac{10}{(4-1)(3-1)}$$

$$= \frac{10}{6}$$

$$= 1.67$$

Source of Variation	Sum of squares	df	Mean squares	F-ratio
Columns	18	(4-1) 3	6	$F_c = \frac{MSC}{MSE} = \frac{6}{1.67} = 3.59$
Rows	8	(3-1) 2	4	
Error	10	(4-1)(3-1) 6	1.67	$F_r = \frac{MSR}{MSE} = \frac{4}{1.67} = 2.39$
Total	36	11		

Columns

$$df = (c-1), (c-1)(r-1)$$

$$= (4-1), (3-1)(4-1)$$

$$= 3, 6$$

$$F_{cr} = 4.76$$

$$F_{cal} < F_{cr}$$

Hypothesis Accepted

Rows

$$df = (r-1), (c-1)(r-1)$$

$$= (3-1), (4-1)(3-1)$$

$$= 2, 6$$

$$F_{cr} = 5.14$$

$$F_{cal} < F_{cr}$$

Hypothesis Accepted

2016 case study

2) The following data represents the sales (RS '000) per month of four brands of soaps allocated among four cities

Brand	city			
	A	B	C	D
I	12	48	30	29
II	42	57	54	39
III	19	42	21	36
IV	28	39	44	50

Test whether

i) The mean sales of three brands are equal

ii) The mean sales of the soaps in each city are equal

Sol

Brand	city				
	A	B	C	D	
I	12	48	30	29	— 119
II	42	57	54	39	— 192
III	19	42	21	36	— 118
IV	28	39	44	50	— 161
	1	1	1	1	590
	101	186	149	154	

C-Ho: Sales of four brands are equal

R-Ho: Sales of four soaps are equal in each city

T = Sum of all observations

$$T = 590$$

N = No. of observation

$$N = 16$$

$$\begin{aligned}\text{ss correction factor} &= \frac{T^2}{N} \\ &= \frac{590^2}{16} = \frac{348100}{16} \\ &= 21756.25\end{aligned}$$

$$\begin{aligned}\text{SST} &= \text{Sum of square of all observations} - \text{correction factor } \left(\frac{T^2}{N}\right) \\ &= (12^2 + 4^2 + 19^2 + 28^2 + \dots + 50^2) - 21756.25 \\ &= 24242 - 21756.25\end{aligned}$$

$$\text{SST} = 2485.75$$

$$\begin{aligned}\text{SSC} &= \frac{(\sum x_1)^2}{N_1} + \frac{(\sum x_2)^2}{N_2} + \frac{(\sum x_3)^2}{N_3} + \frac{(\sum x_4)^2}{N_4} - \frac{T^2}{N} \\ &= \frac{101^2}{4} + \frac{186^2}{4} + \frac{149^2}{4} + \frac{154^2}{4} - 21756.25 \\ &= 22678.5 - 21756.25\end{aligned}$$

$$\text{SSC} = 922.25$$

$$\begin{aligned}\text{SSR} &= \frac{(\sum x_{ji})^2}{N_1} + \frac{(\sum x_{j2})^2}{N_2} + \frac{(\sum x_{j3})^2}{N_3} + \frac{(\sum x_{j4})^2}{N_4} - \frac{T^2}{N} \\ &= \frac{119^2}{4} + \frac{118^2}{4} + \frac{161^2}{4} + \frac{192^2}{4} - 21756.25 \\ &= 22717.5 - 21756.25\end{aligned}$$

$$\text{SSR} = 961.25$$

$$SSE = SSC - SST - SSR$$

$$= 2485.75 - 922.25 - 961.25$$

$$SSE = 602.25$$

$$MSC = \frac{SSC}{C-1} = \frac{922.25}{4-1} = 307.4167$$

$$MSC = 307.42$$

$$MSR = \frac{SSR}{r-1} = \frac{961.25}{4-1}$$

$$MSR = 320.42$$

$$MSE = \frac{SSE}{(r-1)(c-1)}$$

$$= \frac{602.25}{(4-1)(4-1)} = \frac{602.25}{9}$$

$$MSE = 66.92$$

Source of variation	Sum of Squares	df	Mean Square	F ratio
columns	922.25	3	307.42	$F_c = \frac{MSC}{MSE} = \frac{307.42}{66.92}$
Rows	961.25	3	320.42	
Error	602.25	9	66.92	$F_c = 4.5938$
Total	2485.75	15		$F_r = \frac{MSR}{MSE} = \frac{320.42}{66.92}$
				$F_r = 4.788$

column

$$df = (c-1), (c-1)(r-1)$$

$$= (4-1), (4-1)(4-1)$$

$$= 3, 9$$

$$F_{cr} = 3.86$$

$$F_{cal} > F_{cr}$$

Hypothesis rejected

Rows

$$df = (r-1), (c-1)(r-1)$$

$$= (4-1), (4-1)(4-1)$$

$$= 3, 9$$

$$F_{cr} = 3.86$$

$$F_{cal} > F_{cr}$$

Hypothesis Rejected.

2015 case study

3) four experiments determine the moisture content of samples of a powder, taking a sample from each of the six consignments. Their assessments are given below

observers	consignments					
	1	2	3	4	5	6
1	9	10	9	10	11	11
2	12	11	9	11	10	10
3	11	10	10	12	11	10
4	12	13	11	14	12	10

Analyse the data and discuss whether there is any significant difference between the consignments or between the observers.

2014 case study

4) You are given the following data relating to the number of table fans sold by four sales men in 3 months April, May, June

Months	Salesman			
	A	B	C	D
April	50	40	48	39
May	46	48	50	45
June	39	44	40	39

i) Is there significant difference in the sales made by the four salesmen

ii) Is there a significant difference in the sales made during different months.

Saturday
16/12/19

Sign Test

One of the easiest Non-Parametric Test is the Sign Test which is based on the direction of plus or minus sign of observations in a Sample instead of their numerical values. There are two types of Sign Test

1. One Sample Sign Test.
2. Two Sample Sign Test.

One Sample Sign Test:-

- * It is a simple non-Parametric Test applicable on the assumption that a population having continuous symmetrical distribution is considered.
- * Therefore the Probability of getting a value less than the mean is 0.5. Similarly the Probability of getting a value greater than the mean is also 0.5.
- * To Test the Null Hypothesis $\mu = \mu_0$ against an appropriate alternative Hypothesis, each Sample value greater than μ_0 is replaced by plus (+) Sign and each Sample value less than μ_0 is replaced by minus (-) Sign. If a Sample value happen to be equal to μ_0 , it is ignored.
- * To Perform the Sign Test, there two methods
 - a) When the Sample is small, the Test is performed by computing Binomial Probabilities by Referring to Binomial probabilities table.
 - b) When the Sample is large, the Test is performed by Normal distribution used as an approximation of the Binomial distribution.
- ▷ Test the Hypothesis that the mean value μ of a continuous distribution is 20 against the alternative Hypothesis $\mu \neq 20$. 15 observations were taken and the following results were obtained
18, 19, 25, 21, 16, 15, 19, 22, 24, 21, 18, 17, 15, 26, 24. Use 0.50 level of Significance.

Sol Null Hypothesis $H_0: P=0.5$

Alternative Hypothesis $H_1: P \neq 0.5$

Replacing each value greater than 20 with (+) plus sign and each value less than 20 with (-) minus sign.

18	19	25	21	16	15	19	22	24	21	18	17	15	26	24
-	-	+	+	-	-	-	+	+	+	-	-	-	+	+

No. of plus signs = 7

No. of minus signs = 8

To check whether 7 plus signs in 15 trials support the Null hypothesis $P=0.5$ or the alternative hypothesis then $P \neq 0.5$.

Find the probability of 7 or more successes.

Using Appendix Table 3

$$= 0.1964 + 0.1964 + 0.1527 + 0.0916 + 0.0417 + 0.0139 + 0.0032 + 0.0005$$

$$= 0.6964$$

$$0.6964 > 0.5$$

Accept the Null Hypothesis.

Two Sample Sign Test

* The Sign Test can be applied to problems that deal with paired data (2 samples).

* In such problems, each pair must be replaced with a plus (+) sign if the first value is greater than the second or a minus (-) sign if the first value is smaller than the second. Suppose the two values in the pair are equal, they are ignored.

1) The following table shows the ratings assigned to two brands of cool drinks X and Y by 12 consumers. Each respondent was asked to taste the two brands of cool drinks and then rate them

Brand X	26	30	44	23	18	50	34	16	25	49	37	20
Brand Y	22	27	39	7	11	56	30	14	18	51	33	16

Sol Null Hypothesis $H_0: P = (0.5) \frac{1}{2}$

Alternative Hypothesis $H_1: P \neq \frac{1}{2}$

26	30	44	23	18	50	34	16	25	49	37	20
22	27	39	7	11	56	30	14	18	51	33	16
+	+	+	+	+	-	+	+	+	-	+	+

No. of plus signs = 10

No. of minus signs = 2

Now to test whether 10 in 12 trials will support the Null Hypothesis

$P = \frac{1}{2}$ or the alternative hypothesis $P \neq \frac{1}{2}$ or $P > \frac{1}{2}$.

The Null Hypothesis H_0 implies that both the brands enjoy equal preferences and none is better than the other. The alternative Hypothesis H_1 means that brand X is Better than brand Y.

Now from Appendix Table 3

The probability of 10 or more successes in 12 trials.

$$= 0.0161 + 0.0029 + 0.0002$$

$$= 0.0192$$

We observe $0.0192 < 0.5$

Hence Null hypothesis is Rejected.

It means $P > \frac{1}{2}$ is accepted.

That is Brand X is better than Brand Y.

Qy The following data pertains to down times (periods in which the computers were inoperative on account of failures, in minutes) of two different computers.

Computer A	58	60	42	62	65	59	60	52	50	75	59	32	48	50	41	45	40	43	48	70	60	80
Computer B	52	57	30	46	66	40	78	55	52	58	44	45	36	56	40	70	50	53	50	30	42	45

Test the null hypothesis that the average downtime is same for

both the computers A and B.

Sol $H_0: P = 1/2$

$H_1: P \neq 1/2$

Computer A	58	60	42	62	65	59	60	52	50	75	59	32	48	50	41	45	40	43	43	70	60	80
Computer B	52	57	30	46	66	40	78	55	52	58	44	45	36	56	40	70	50	53	50	30	42	45
Signs	+	+	+	+	-	+	-	-	-	+	+	-	+	-	+	-	-	-	-	+	+	+

No. of plus signs = 12

No. of minus signs = 10

Now we have to test whether 12 successes in 22 trials support the null hypothesis $P = 1/2$.

Since the number of trials $n = 22$, the Binomial probabilities cannot be used, hence it is approximated to Normal distribution

$$Z = \frac{x - \mu}{\sigma}$$
$$= \frac{x - np}{\sqrt{npq}}$$

$$= \frac{12 - (22 \times 1/2)}{\sqrt{22 \times 1/2 \times 1/2}}$$

$$= \frac{1}{\sqrt{5.5}} \Rightarrow \frac{1}{2.34}$$

$$= 0.43$$

The table value of Z at 0.50 level of significance lies between -1.96 and +1.96

We observe that the calculated value 0.43 is less than the critical value of Z at 0.50 level of significance. Hence

Null Hypothesis is Accepted

i.e., the downtimes of computer A and computer B are equal.

Sunday
7/12/19

Wilcoxon Matched pair test (Signed Rank Test)

This test not only Tests the direction, but also the magnitude of differences between matched pairs like husband and wife, twins, output of two similar machines etc.

Procedure:-

- Step ①:- Find the difference 'd' between each pairs of scores.
- Step ②:- Rank the 'd's without regard to sign, giving 1st first rank to the smallest difference, Second Rank to the next in order and so on.
- Step ③:- Then to each rank Affix Sign of difference.
- Step ④:- Total ^{the} ranks with plus (+) Signs and total ranks with minus (-) Signs.
- Step ⑤:- The pairs whose Scores are equal are dropped from the calculation. When two or more pairs have the same 'd' value, their ranks are averaged.
- Step ⑥:- The Sum of the ranks of the smaller sign group is designated as 'T'. This value is compared with the table value and accept and Reject criteria are decided.

1) The following data refers to the consumer awareness of the bank services offered before and after an Advertisement campaigning

Consumer	1	2	3	4	5	6	7	8	9	10
Before Ad campaign	82	81	89	74	68	80	77	66	77	75
After Ad campaign	87	84	84	76	78	81	79	81	81	83

use wilcoxon matched pair test, Test the hypothesis that there is no difference in consumer awareness of Bank Services offered after the advertisement campaigning.

Sol H_0 : There is no difference in consumer awareness of bank services after Ad campaign.

H_1 : There is difference in consumer awareness of bank services after Ad campaign.

Consumer	After Ad campaign	Before Ad Campaign	difference d _i	Rank of d _i	Rank (-) Sign	Rank (+) Sign
1	87	82	5	6.5		6.5
2	84	81	3	4		4
3	84	89	-5	6.5	-6.5	-
4	76	74	2	2.5		2.5
5	78	68	10	9		9
6	81	80	1	1		1
7	79	77	2	2.5		2.5
8	81	66	15	10		10
9	81	77	4	5		5
10	83	75	8	8		8
					-6.5	48.5

$T =$ computed sum of Ranks of Smaller Sign group = 6.5

Now this 'T' value is compared with the critical value of 'T' in wilcoxon matched pair Test from Table 8 for

$$T_{crit} = 8$$

we observed $T_{cal} < T_{crit}$

Reject Null Hypothesis i.e., There is a difference between the consumer Awareness before and after the Ad campaign.

The MANN - WHITNEY Test - U Test

This test belongs to rank the Sum Test. It uses Ranking Information rather than pluses and Minuses. The U Test is designed to determine whether two independent Samples have been drawn from the same population or from two different populations having the same distribution.

To apply U-Test to a set of two Samples (Ex:- The average grades of students of two different colleges), the values

of the combined samples n_1 and n_2 are ranked from the lowest to the highest Rank. Irrespective of the samples, Rank one (1) to the lowest score, Rank two (2) to the next in order and soon. Indicating each the identification of the sample. Then the Ranks of each sample group are summed separately and the sums are represented as R_1 and R_2 . Then the value of ~~Mann~~ U-statistic is computed using the following formula.

$$U = n_1 n_2 + \left[\frac{n_1(n_1+1)}{2} \right] - R_1$$

Where n_1 = number of items in Sample 1.

n_2 = Number of items in Sample 2.

R_1 = Sum of Ranks of items in Sample 1.

Refer Table - 9 and find U-value at n_1 and n_2 at 0.05 or 0.01 level of Significance.

$U_{cal} > U_{crit}$ — Accept the Null Hypothesis.

1) The following data indicates the lifetime (in hours) of samples of two kinds of light bulbs in continuous use using Mann-Whitney Test to compare the life times of two brands A and B light bulbs are same.

Sol Brand A 603 625 641 622 585 593 660 600 633 580 615
Brand B 620 640 646 620 652 639 590 646 631 669 610

$\frac{645}{619}$

6+ 13+ 18+
12+ 2+ 4
23+ 5+
15+ 1+
7+ 10

Sol

H_0 : The mean lifetime of the two brands bulbs A and B are equal.

H_1 : The mean lifetime of the two brands bulbs A and B are not equal.

Sample	Group	Rank
580	A	1
585	A	2
590	B	3
593	A	4
600	A	5
603	A	6
610	B	7
615	A	8
619	B	9
620	B	10.5
620	B	10.5
622	A	12
625	A	13
631	B	14
633	A	15
639	B	16
640	B	17
641	A	18
645	A	19
646	B	20.5
646	B	20.5
652	B	22
660	A	23
669	B	24

$$U_{cal} = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$n_1 = 12 \quad n_2 = 12$$

$$R_1 = 1 + 2 + 4 + 5 + 6 + 8 + 12 + 13 + 15 + 18 + 19 + 23$$

$$R_1 = 126$$

$$U = (12 \times 12) + \frac{12(12+1)}{2} - 126$$

$$= 144 + 78 - 126$$

$$= 222 - 126$$

$$U_{cal} = 96$$

from Table 9 for n_1 & n_2 and at 5% level of significance

$$U_{cr} = 37$$

We observe

$$U_{cal} (96) > U_{cr} (37)$$

Accept the Null Hypothesis.

$$\frac{20+21}{2} = \frac{41}{2} = 20.5$$

Kolmogorov - Smirnov Test (K-S)

The K-S Test is a test of goodness of fit in which the cumulative frequency distribution is considered which would occur under the theoretical distribution and compared that with observed cumulative frequency distribution.

Steps :-

1. Develop cumulative frequency distribution from the given sample.
2. Determine the largest absolute deviation between the distribution
3. This value is compared with critical (Table) value to test its significance.
4. The Theoretical Distribution represents our expectations under H_0 .
5. The point of greatest divergence between the observed and Theoretical distribution is determined and this value is identified as 'D' (Maximum Deviation).

$$D = \text{Max} | F_o(x) - F_t(x) |$$

This calculated value of 'D' is compared with Table value of D from Table 10 and Decision on Significance is made.

$$KS - D_{\text{cal}} > KS - D_{\alpha; i} \quad \text{Reject } H_0$$

1) A company has conducted a field Survey covering 200 students. Apart from other questions, it asked the respondents to indicate on a 5-point scale on how much the durability of a particular product is important to them. The response indicated as follows.

Very important - 50
Some what important - 60
Neither imp nor unimportant - 20
Some what unimportant - 40
Very unimportant - 30

200

Using K-S Test, Test the hypothesis that there is no difference in importance of the ratings for durability among the respondents.

Sol H_0 : There is no difference in the importance of ratings for durability among the respondents.

H_1 : There is a difference in the importance of ratings for durability among the respondents.

Worksheet

Importance of Durability	Observed Number	Observed Proportion	Observed Cumulative Proportion	Null Proportion	Null Cumulative Proportion	Absolute diffn of observed Null
Very imp	50	$50/200 = 0.25$	0.25	$\frac{200}{1000} = 0.2$	0.2	$(0.25 - 0.2)$ 0.05
Somewhat imp	60	$60/200 = 0.30$	0.55	$\frac{200}{1000} = 0.2$	0.4	$(0.55 - 0.4)$ <u>0.15</u>
Neither imp nor unimp	20	$20/200 = 0.10$	0.65	0.2	0.6	0.05
Somewhat unimp	40	$40/200 = 0.20$	0.85	0.2	0.8	0.05
Very unimp	30	$30/200 = 0.15$	1.00	0.2	1.0	0

from the above table, the largest absolute difference is 0.15 which is $KS-D_{cal} = 0.15$

Referring to Table 10 for critical values for a sample size of 200, the critical value of 'D' at 0.05 level of significance is $\frac{1.36}{\sqrt{n}}$

$$\Rightarrow \frac{1.36}{\sqrt{200}} = \frac{1.36}{14.14}$$

$$\Rightarrow 0.0962$$

We $KS-D_{cal} (0.15) > KS-D_{cri} (0.0962)$

Hence the Null Hypothesis is Rejected.

Kruskal Wallis Test

This Test is a direct generalisation of Mann-Whitney Test where there exist three or more independent groups. It tests the Null Hypothesis that all samples come from identical populations and the alternative Hypothesis is that the means of all the populations are not equal. To perform ^{the} Kruskal Wallis Test, Rank all the scores without regard to groups to which they belong and then compute the Sum of the Ranks for each group. The Sums are denoted by R_i . If the Null hypothesis is true, we can expect R_i 's to be more or less equal. The Test statistic in Kruskal Wallis Test

$$H = \frac{12}{n(n+1)} \sum_{i=1}^K \frac{R_i^2}{n_i} - 3(n+1).$$

where $n = n_1 + n_2 + n_3 + \dots + n_K$ is the total number of observations in all samples and R_i is the Sum of Ranks of all items in sample 'i'.

If the Null Hypothesis is true and each sample is at least of size 5, the sampling distribution of this statistic can be approximated closely with χ^2 distribution with $K-1$ df and the calculated value is compared with chi-square table value at 5% level of significance and if $H_{cal} > H_{tab}$

Monday
18/12/17

$H_{cal} > \chi_{0.05}^2 - \text{Accepted.}$

1) Three groups of Salesman of a company have undergone training. The method of Training used was different for each group. When the training was completed, the salesman were given a test the marks scored by them are as follows

Training Method A: 75 83 68 85 90 61

Training Method B: 62 70 67 82 80 87 64

Training Method C: 65 71 74 63 89

Using Kruskal Wallis Test, find out whether there is a difference in the effectiveness of the three training methods.

Sol H_0 : There is no difference in the effectiveness of three training methods.

H_1 : There is a difference in the effectiveness of three training methods.

Marks	Group	Rank
90	A	1
89	C	2
87	B	3
85	A	4
83	A	5
82	B	6
80	B	7
75	A	8
74	C	9
71	C	10
70	B	11
68	A	12
67	B	13
65	C	14
64	B	15
63	C	16
62	B	17
61	A	18

Ranks in Sample (1) $\Rightarrow 1, 4, 5, 8, 12, 18$

$$R_1 = 1 + 4 + 5 + 8 + 12 + 18$$

$$R_1 = 48$$

Ranks in Sample (2) $\Rightarrow 3, 6, 7, 11, 13, 15, 17$

$$R_2 = 3 + 6 + 7 + 11 + 13 + 15 + 17$$

$$R_2 = 72$$

Ranks in Sample (3) $\Rightarrow 2, 9, 10, 14, 16$

$$R_3 = 2 + 9 + 10 + 14 + 16$$

$$R_3 = 51$$

Given $n_1 = 6, n_2 = 7, n_3 = 5$

$$n = n_1 + n_2 + n_3$$

$$n = 6 + 7 + 5$$

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{18(18+1)} \sum_{i=1}^3 \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{342} \left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} \right] - 3(18+1)$$

$$= 0.035 \left[\frac{48^2}{6} + \frac{72^2}{7} + \frac{51^2}{5} \right] - 57$$

$$= 0.035 \left[\frac{2304}{6} + \frac{5184}{7} + \frac{2601}{5} \right] - 57$$

$$= 0.035 [384 + 740.57 + 520.2] - 57$$

$$= 0.035 (1644.7) - 57$$

$$= 57.57 - 57$$

$$= 0.57$$

Using chisquare distribution we have $df = 3 - 1 = 2$.

Level of significance = 0.05 $\chi^2_{0.05/2} = 5.991$, $\chi^2_{cal} > \chi^2_{crit}$ Rejected.

Hence the Null hypothesis is rejected it means there is a difference between the Training Methods A, B and C given for the salesman

Test	calculator	Comparing	Formula / Table	Accept/Reject H_0
Sign Test Table 3	Small Sample - Binomial Distribution Large Sample - Normal Distribution	> 0.5 $-1.96 \leq Z \leq 1.96$	$Z = \frac{x - 0.5n}{\sigma}$	> 0.5 - Accept H_0 Z_{cal} lies b/w ± 1.96 - Accept H_0
Median Test Table 5				
Sign Ranked Test Wilcoxon paired Table 8	T_{cal} = Sum of Ranks of Smaller Sign group	T_{cri} for n values	Table 8	$T_{cal} < T_{cri}$ Reject H_0
Mann-Whitney Table 9	$U = n_1 n_2 + \left[\frac{n_1(n_1+1)}{2} \right] - R_1$	U_{cri}	Table 9	$U_{cal} > U_{cri}$ Accept H_0
Kolmogorov Smirnov Table 10	Largest Absolute deviation $D_{cal} = F_0(x) - F_t(x) $	$KS - D_{cri}$	Table 10	$KS - D_{cal} > KS - D_{cri}$ Reject H_0
Kruskal Wallis Test Table 5	$H = \frac{12}{n(n+1)} \sum_{i=1}^K \frac{R_i^2}{n_i} - 3(n+1)$	$\chi^2_{0.05, (k-1) df}$	Table 5	$H_{cal} > \chi^2_{0.05, (k-1) df}$ Accept H_0 .

Tuesday
19/12/17

MEDIAN TEST [Two sample or K sample median test]

The median Test determines the significance of the difference between the medians of two or more random samples. The procedure is as follows

- Step ①: - Find the median of two or more samples combined.
- Step ②: Count the Number of values in each sample falling above the combined median and the items that are not above the median and prepare a contingent table.
- Step ③: Perform a chi-square test for the above median cases and not above median cases so counted.
- Step ④: Determine the Degrees of freedom df.
- Step ⑤: Compare the calculated value of chi-square with the corresponding table value at specified degrees of freedom and at specific level of Significance.
- Step ⑥: Make the Decision that is if the calculated value equals or Exceeds the Table value, Reject the Null Hypothesis.

1) The following data relates to the down times of two computers perform the median test.

Computer A 58 60 42 62 65 59 60 52 50 75 59
 52 57 30 46 66 40 78 55 52 58 44

Computer B 32 48 50 41 45 40 43 43 70 60 80
 45 36 56 40 70 50 53 50 30 42 45

Sol

Median for row one = 59
Median for row Two = 40

Arranging the numbers for Median calculation.

Row 1: 42 50 52 58 59 (59) 60 60 62 65 75

Row 2: 30 40 44 46 52 (52) 55 57 58 66 78

Row 3: 32 40 41 43 43 (45) 48 50 60 70 80

Row 4: 30 36 40 42 45 (45) 50 50 53 56 70

$$\text{Combined Median} = \frac{89+52+45+45}{4} = 50.25 \sim 50$$

In ~~re A~~ A the number of values below 50 = 5
 computer the number of values above 50 = 17

In ~~re B~~ B the number of values below 50 = 13
 computer the number of values above 50 = 9

H_0 : There is no difference in the medians of downtimes for the two computers.

H_1 : There is a difference in the medians of downtimes for the two computers.

calculation of χ^2 value

	Below Median	Above median	Total
Computer A	5	17	22
Computer B	13	9	22
Total	18	26	44

Observed frequency	Expected frequency	$O_i - e_i$	$(O_i - e_i)^2$	$\frac{(O_i - e_i)^2}{e_i}$	
5	11	-6	36	3.27	$e_i = \frac{44}{4}$ $= 11$
17	11	6	36	3.27	
13	11	2	4	0.36	
9	11	-2	4	0.36	

$$\chi^2_{\text{cal}} = \sum \frac{(O_i - e_i)^2}{e_i}$$

$$= 7.26$$

from Tables

$$df = (c-1)(r-1) \Rightarrow df = (2-1)(2-1) = 1$$

At 0.05 level of Significance $\chi^2_{0.05,1} = 3.841$

$\chi^2_{\text{cal}} > \chi^2_{0.05,1}$ Reject the Null Hypothesis.

Hence there is a difference in the medians of the downtimes of the two computers.