KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University Established Under Section 3 of UGC Act 1956)

Pollachi Main Road, Eachanari (Post)

Coimbatore –641 021

DEPARTMENT OF MATHEMATICS

17BAU102 CORE-STATISTICS FOR BUSINESS DECISIONS

Semester I

L	T	P	C
5	-	-	5

……

OBJECTIVES: To familiarize the students with various Statistical Data Analysis tools that can be used for effective decision making. Emphasis will be on the application of the concepts learnt.

UNIT I

Measures of Central Value: Characteristics of an ideal measure; Measures of Central Tendency - mean, median, mode, harmonic mean and geometric mean. Merits, Limitations and Suitability of averages. Relationship between averages. Measures of Dispersion: Meaning and Significance. Absolute and Relative measures of dispersion - Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation, Moments, Skewness, Kurtosis.

UNIT II

Correlation Analysis: Meaning and significance. Correlation and Causation, Types of correlation. Methods of studying simple correlation - Scatter diagram, Karl Pearson's coefficient of correlation, Spearman's Rank correlation coefficient, Regression Analysis: Meaning and significance, Regression vs. Correlation. Linear Regression, Regression lines (X on Y, Y on X) and Standard error of estimate.

UNIT III

Analysis of Time Series: Meaning and significance. Utility, Components of time series, Models (Additive and Multiplicative), Measurement of trend: Method of least squares, Parabolic trend and logarithmic trend.

UNIT IV

Index Numbers: Meaning and significance, problems in construction of index numbers, methods of constructing index numbers-weighted and unweighted, Test of adequacy of index numbers, chain index numbers, base shifting, splicing and deflating index number.

UNIT V

Probability: Meaning and need. Theorems of addition and multiplication. Conditional probability. Bayes' theorem, Random Variable- discrete and continuous. Probability Distribution: Meaning, characteristics (Expectation and variance) of Binomial, Poisson, and Normal distribution. Central limit theorem.

SUGGESTED READINGS:

TEXT BOOKS

1. Gupta, S.P. Statistical Methods (34th ed.). New Delhi: Sultan Chand & Sons.

2. Richard Levin & David Rubin . Statistics for management. New Delhi: Prentice Hall.

3. Anderson, Sweeny, & Williams. Statistics for Business and Economics. South Western.

REFERENCES

1. Navnitham , P.A .(2004). Business Mathematics and Statistics. Trichy: Jai Publications.

2. Pillai, R.S.N., & Bagavathi , V. (2002). Statistics . New Delhi: S. Chand & Company Ltd

3. Srivastava, T N., & Shailaja Rego. (2012). Statistics for Management. New Delhi: Mc Graw Hill Education .

4. Amir, D., Aczel & Jayavel Sounderpandian. (2012). Complete Business Statistics (7th ed.). New Delhi: Mc Graw Hill Education.

5. Dr. Arora, P.N. (1997). A foundation course statistics. New Delhi: S.chand & Company Ltd.

…….

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Post)
Coimbatore –641 021

DEPARTMENT OF MATHEMATICS

SUBJECT: STATISTICS FOR BUSINESS DECISIONS SEMESTER I L T P C
SUBJECT CODE:17BAU102 5 0 0 5

S.No	Lecture Duration	Topic to be covered	Support Material
		UNIT – I	
1	1	Measures of central value-Introduction-Characteristics of an Ideal measure	T1:Chap-7;Pg:178-183
2	1	Continuation of Measures of central value-Introduction-Characteristics of an Ideal measure	T1:Chap-7;Pg:183-186
3	1	Measure of Central Tendency :Mean	T1:Chap-7;Pg:186-188
4	1	Measure of Central Tendency – Median problems	R2:Chap-9;pg:146-150
5	1	Continuation of Measure of Central Tendency – Median problems	R2:Chap-9;pg:150-156
6	1	Mode –Problems	T1:Chap-7;Pg:211-215
7	1	Continuation of mode problems	T1:Chap-7;Pg:215-220
8	1	Harmonic mean	T1:Chap-7;Pg:222-225
9	1	Continuation of Harmonic mean	T1:Chap-7;Pg:226-229
10	1	Geometric mean	T1:Chap-7;Pg:229-234
11	1	Continuation of Geometric mean	T1:Chap-7;Pg:235 -238
12	1	Merits, Limitations and suitability of averages. Relationship between averages	R2:Chap-2;Pg:58-59
13	1	Measures of dispersion- Introduction- Absolute & relative measures of dispersion	T1:Chap-8;Pg:268-271
14	1	Range, Quartile deviation	T1:Chap-8;Pg:271-274
15	1	Continuation of Range, Quartile deviation	T1:Chap-8;Pg:274-277
16	1	Mean deviation and Standard deviation	T1:Chap-8;Pg:277-282
17	1	Continuation of Mean deviation and Standard deviation	T1:Chap-8; pg:282-291
18	1	Coefficient of variation	T1:Chap-8;Pg:293-295
19	1	Coefficient of variation	T1:Chap-8;Pg:295-297
20	1	Skewness ,Moments and Kurtosis	R1:Chap-8;Pg:245-253
21	1	Continuation of Skewness ,Moments and	R1:Chap-8; pg:253-259

		Kurtosis	
22	**1**	**Recapitulation and discussion of possible questions.**	
		Total – 22 Hours	
		T1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons **R1.Navnitham .P.A .(2004).Business Mathematics and statistics.Trichy:Jai Publications.** **R2.pillai,R.S.N.,& Bagavathi , v.(2002),statistics.New Delhi:S.Chand &Company Ltd**	

		UNIT –II	
1	**1**	**Correlation Analysis- Introduction, Correlation and causation**	**T1:Chap-10;Pg:378-381**
2	**1**	**Types of correlation – Scatter diagram**	**T1:Chap-10;pg:381-385**
3	**1**	**Karl Pearson's coefficient of correlation**	**R1:Chap-11;Pg:318-321**
4	**1**	**Continuation of Karl Pearson's coefficient of correlation**	**R1:Chap-11;Pg:322-324**
5	**1**	**Spearman's rank correlation coefficient**	**R5:Chap-4;Pg:125-128**
6	**1**	**Continuation of Spearman's rank correlation coefficient**	**R5:Chap-4;pg:128 - 132**
7	**1**	**Regression Analysis-Introduction**	**T1:Chap-11;Pg:436-437**
8	**1**	**Linear Regression, Regression lines**	**R5:Chap-5;Pg:150-151**
9	**1**	**Properties of Regression lines**	**R5:Chap-4;Pg:152-153**
10	**1**	**Problems on Regression lines** **Standard error of estimate**	**T1:Chap-11;Pg:442-447** **T1:Chap-11;Pg:438**
11	**1**	**Continuation of Standard error of estimate**	**T1:Chap-11;Pg:453-454**
12	**1**	**Recapitulation and discussion of possible questions.**	
		Total – 12 Hours	

		T1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons R1.Navnitham .P.A .(2004).Business Mathematics and statistics.Trichy:Jai Publications. R5.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand &Company Ltd.	

		UNIT - III	
1	1	Analysis of Time series- Utility and Components	T1:Chap-14;Pg:590-595
2	1	Continuation of Analysis of Time series- Utility and Components	T1:Chap-14;Pg:596-599
3	1	Models(Additive and Multiplicative)	R5:Chap-14;Pg:495-496
4	1	Method of least squares - Problems	T1:Chap-14;Pg:613-614
5	1	Continuation of Problems related to Method of least squares	T1:Chap-14;Pg:614-619
6	1	Parabolic trend - Problems	T1:Chap-14;Pg:619-620
7	1	Continuation of Problems related to parabolic trend	T1:Chap-14;Pg:620-622
8	1	Logarithmic trend	T1:Chap-14;Pg:622-624
9	1	Continuation of Logarithmic trend	T1:Chap-14;Pg:623-626
10	1	Recapitulation and discussion of possible questions.	
		TOTAL – 10 Hours	
		T1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons R5.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand &Company Ltd.	

		UNIT – IV	
1	1	Index number-Introduction-Problems in the construction of index number	R3:Chap-16;Pg:922-924
2	1	Methods of constructing index number	R2:Chap-14;Pg:483-484
3	1	Un weighted index number - Problems	R5:Chap-13;Pg:449-451
4	1	Continuation of Un weighted index number - Problems	R5:Chap-13;Pg:451-454
5	1	Weighted index number - Problems	R5:Chap-13;Pg:456-458
6	1	Continuation of Weighted index number - Problems	R5:Chap-13;Pg:458-460
7	1	Continuation of Problems to Weighted index number	R5:Chap-13;Pg:461-465
8	1	Test of adequacy of index number	T1:Chap-13;Pg:539-542
9	1	Continuation of Test of adequacy of index	T1:Chap-13;Pg:542-545

		number	
10	**1**	**Base shifting, Splicing and Deflating index number**	**T1:Chap-13;Pg:545-548**
11	**1**	**Deflating index number- Problems**	**T1:Chap-13;Pg:548-551**
12	**1**	**Continuation of Deflating index number- Problems**	**T1:Chap-13;Pg:551-553**
13	**1**	**Recapitulation and discussion of possible questions.**	
		Total – 13 hours	
		T1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons **R2.pillai,R.S.N.,& Bagavathi , v.(2002),statistics.New Delhi:S.Chand &Company Ltd** **R3.Srivastava ,T N., & Shailaja Rego .(2012).statistics for management .New Delhi:Mc Graw Hill Education.** **R5.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand &Company Ltd.**	

		UNIT –V	
1	**1**	**Probability-Introduction-Basic definition and problems**	**T1:Chap-1-vol-I;Pg:752-757**
2	**1**	**Continuation of Probability-Introduction-Basic definition and problems**	**T1:Chap-1-vol-I;Pg:758-760**
3	**1**	**Theorem of addition and multiplication, Conditional Probability and Baye's theorem**	**T1:Chap-1;Pg:761-765**
4	**1**	**Continuation of Theorem of addition and multiplication, Conditional Probability and Baye's theorem**	**T1:Chap-1;pg:765-771**
5	**1**	**Random variable: discrete and continuous of Probability distribution**	**R5:Chap-6;Pg:244-248**
6	**1**	**Continuation of Random variable: discrete and continuous of Probability distribution**	**R5:Chap-6;Pg:249-254**
7	**1**	**Problems on probability distribution Binomial distribution**	**R5:Chap-6;Pg:250-256**
8	**1**	**Continuation of Problems on probability distribution Binomial distribution**	**R2:Chap-18;pg:723-728**
9	**1**	**Problems on poison distribution**	**T1:Chap-1;Pg:817-824**
10	**1**	**Continuation problems of Poisson distribution**	**R5:Chap-6;Pg:271-276**
11	**1**	**Problems on Poisson distribution**	**T1:Chap-1;Pg:831-834**

		Normal distribution	**R5:Chap-6;pg:285-288**
12	**1**	**Problems on normal distribution**	**R5:Chap-6;Pg:289-294**
13	**1**	**Continuation of Problems on normal distribution**	**R5:Chap-6;Pg:294-299**
14	**1**	**Central limit theorem.**	**T2:Chap6:pg:319-323**
15	**1**	**Recapitulation and discussion of possible questions.**	
16	**1**	**Discussion of previous ESE question papers**	
17	**1**	**Discussion of previous ESE question papers**	
18	**1**	**Discussion of previous ESE question papers**	
		Total -18hours	
		T1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons **T2. Richard Levin & David Rubin. Statistic for management, New Delhi: Prentice Hall** **R2.pillai,R.S.N.,& Bagavathi , v.(2002),statistics.New Delhi:S.Chand &Company Ltd** **R5.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand &Company Ltd**	

TEXTBOOK:

T1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons

T2. Richard Levin & David Rubin. Statistic for management, New Delhi: Prentice Hall

T3. Andreson, Sweeny, &Williams. Statistics for Bussines and Economics. South Western.

REFERENCES:

R1.Navnitham .P.A .(2004).Business Mathematics and statistics.Trichy:Jai Publications.

R2.pillai,R.S.N.,& Bagavathi , v.(2002),statistics.New Delhi:S.Chand &Company Ltd

R3.Srivastava ,T N., & Shailaja Rego .(2012).statistics for management .New Delhi:Mc Graw Hill Education.

R4.Amir , D.,Aczel & Jayavel sounderpandian.(2012).complete Business Statistics (7th ed.).New Delhi :Mc Hill Education

R5.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand &Company Ltd.

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Post)
Coimbatore –641 021
DEPARTMENT OF MATHEMATICS

SUBJECT: STATISTICS FOR BUSINESS DECISIONS	**SEMESTER: I**	**L T P C**
SUBJECT CODE: 17BAU102	**CLASS:I UG (BBA)**	**5 0 0 5**

UNIT I

Measures of Central Value: Characteristics of an ideal measure; Measures of Central Tendency - mean, median, mode, harmonic mean and geometric mean. Merits, Limitations and Suitability of averages. Relationship between averages. Measures of Dispersion: Meaning and Significance. Absolute and Relative measures of dispersion - Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation, Moments, Skewness, Kurtosis.

TEXT BOOKS

1. Gupta, S.P. Statistical Methods (34th ed.). New Delhi: Sultan Chand & Sons.

REFERENCES

1. Navnitham , P.A .(2004). Business Mathematics and Statistics. Trichy: Jai Publications.

2. Pillai, R.S.N., & Bagavathi , V. (2002). Statistics . New Delhi: S. Chand & Company Ltd

Introduction

In the modern world of computers and information technology, the importance of statistics is very well recogonised by all the disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date there is no other human walk of life, where statistics cannot be applied.

Arithmetic mean or mean :

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values $x_1, x_2 \ldots x_n$ then the mean, $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i$$

This formula is for the ungrouped or raw data.

Example 1 :

Calculate the mean for 2, 4, 6, 8, 10

Solution:

$$\bar{x} = \frac{2+4+6+8+10}{5}$$

$$= \frac{30}{5} = 6$$

Median :

The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

Ungrouped or Raw data :

Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value .If the number of values are even, median is the mean of middle two values.

By formula

$$\text{Median} = \text{Md} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item.}$$

Example 11:

When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

Solution:

Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the 5^{th} item i.e., 25 is the median

Using formula

$$\text{Md} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item.}$$

$$= \left(\frac{9+1}{2}\right)^{\text{th}} \text{item.}$$

Harmonic mean (H.M) :

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If $x_1, x_2 \ldots x_n$ are n observations,

$$H.M = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

For a frequency distribution

$$= \left(\frac{10}{2}\right)^{th} \text{item}$$

$$= 5^{th} \text{ item}$$

$$= 25$$

Geometric mean :

The geometric mean of a series containing n observations is the n^{th} root of the product of the values. If $x_1, x_2 \ldots x_n$ are observations then

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

$$= (x_1 . x_2 \ldots x_n)^{1/n}$$

$$\log GM = \frac{1}{n} \log(x_1 . x_2 \ldots x_n)$$

$$= \frac{1}{n} (\log x_1 + \log x_2 + \ldots + \log x_n$$

$$= \frac{\sum \log x_i}{n}$$

$$GM = \text{Antilog } \frac{\sum \log x_i}{n}$$

For grouped data

$$GM = \text{Antilog}\left[\frac{\sum f \log x_i}{N}\right]$$

Example 8:

Calculate the geometric mean of the following series of monthly income of a batch of families 180,250,490,1400,1050

x	logx
180	2.2553
250	2.3979
490	2.6902
1400	3.1461
1050	3.0212
	13.5107

$$GM = \text{Antilog}\left[\frac{\sum \log x}{n}\right]$$

$$= \text{Antilog } \frac{13.5107}{5}$$

$$= \text{Antilog } 2.7021 \quad = 503.6$$

RELATIONSHIP BETWEEN AVERAGES

In a symmetrical distribution the three simple averages mean = median = mode. For a moderately asymmetrical distribution, the relationship between them are brought by Prof. Karl Pearson as mode = 3median - 2mean.

Example 34:

If the mean and median of a moderately asymmetrical series are 26.8 and 27.9 respectively, what would be its most probable mode?

Solution:

Using the empirical formula

$$\text{Mode} = 3\text{ median} - 2\text{ mean}$$
$$= 3 \times 27.9 - 2 \times 26.8$$
$$= 30.1$$

MEASURES OF DISPERSION

The various absolute and relative measures of dispersion are listed below.

Absolute measure	Relative measure
1. Range	1.Co-efficient of Range
2.Quartile deviation	2.Co-efficient of Quartile deviation
3.Mean deviation	3. Co-efficient of Mean deviation
4.Standard deviation	4.Co-efficient of variation

Range and coefficient of Range:

Range:

This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

In symbols, Range $= L - S$.

Where L = Largest value.

S = Smallest value.

Co-efficient of Range :

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

Example1:

Find the value of range and its co-efficient for the following data.

7, 9, 6, 8, 11, 10, 4

Solution:

L=11, S = 4.

Range $= L - S = 11 - 4 = 7$

$$\text{Co-efficient of Range} = \frac{L - S}{L + S} = \frac{11 - 4}{11 + 4} = \frac{7}{15} = 0.4667$$

Quartile Deviation and Co efficient of Quartile Deviation :

Quartile Deviation (Q.D) :

Definition: Quartile Deviation is half of the difference between the first and third quartiles. Hence, it is called Semi Inter Quartile Range.

In Symbols, $Q.D = \frac{Q_3 - Q_1}{2}$. Among the quartiles Q_1, Q_2 and Q_3, the range $Q_3 - Q_1$ is called inter quartile range and $\frac{Q_3 - Q_1}{2}$, Semi inter quartile range.

Co-efficient of Quartile Deviation :

$$\text{Co-efficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Mean Deviation and Coefficient of Mean Deviation:

Mean Deviation:

The range and quartile deviation are not based on all observations. They are positional measures of dispersion. They do not show any scatter of the observations from an average. The mean deviation is measure of dispersion based on all items in a distribution.

Coefficient of mean deviation:

$$\text{Coefficient of mean deviation:} = \frac{\text{Mean deviation}}{\text{Mean or Median or Mode}}$$

If the result is desired in percentage, the coefficient of mean deviation

$$= \frac{\text{Mean deviation}}{\text{Mean or Median or Mode}} \times 100$$

Example 6:

Calculate mean deviation from mean and median for the following data:

100,150,200,250,360,490,500,600,671 also calculate co-efficients of M.D.

Solution:

$$\text{Mean} = \bar{\mathbf{x}} = \frac{\sum \mathbf{x}}{\mathbf{n}} = \frac{3321}{9} = 369$$

Now arrange the data in ascending order

100, 150, 200, 250, 360, 490, 500, 600, 671

$$\text{Median} = \text{Value of } \left(\frac{n+1}{2}\right)^{th} \text{ item}$$

$$= \text{Value of } \left(\frac{9+1}{2}\right)^{th} \text{ item}$$

$$= \text{Value of } 5^{th} \text{ item}$$

$$= 360$$

X	$\lvert D \rvert = \lvert x - \bar{x} \rvert$	$\lvert D \rvert = \lvert x - Md \rvert$
100	269	260
150	219	210
200	169	160
250	119	110
360	9	0
490	121	130
500	131	140
600	231	240
671	302	311
3321	1570	1561

$$\text{Co-efficient of M.D} = \frac{M.D}{\bar{x}}$$

$$= \frac{174.44}{369} = 0.47$$

$$\text{M.D from median} = \frac{\sum|D|}{n}$$

$$= \frac{1561}{9} = 173.44$$

$$\text{Co-efficient of M.D.} = \frac{\text{M.D}}{\text{Median}} = \frac{173.44}{360} = 0.48$$

Standard Deviation and Coefficient of variation:

Standard Deviation :

Definition:

It is defined as the positive square-root of the arithmetic mean of the Square of the deviations of the given observation from their arithmetic mean.

The standard deviation is denoted by the Greek letter σ (sigma)

$$\text{Thus } \sigma = \sqrt{\left(\frac{\Sigma x^2}{n}\right)} \text{ or } \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

Example 13:

The daily temperature recorded in a city in Russia in a year is given below.

Temperature C^0	No. of days
-40 to –30	10
-30 to –20	18
-20 to –10	30
-10 to 0	42
0 to 10	65
10 to 20	180
20 to 30	20
	365

Calculate Standard Deviation.

Solution:

Temperature	Mid value (m)	No. of days f	$d = \frac{m-(-5^{n})}{10^{n}}$	fd	fd^2
-40 to -30	-35	10	-3	-30	90
-30 to -20	-25	18	-2	-36	72
-20 to -10	-15	30	-1	-30	30
-10 to - 0	-5	42	0	0	0
0 to 10	5	65	1	65	65
10 to 20	15	180	2	360	720
20 to 30	25	20	3	60	180
		N=365		Σfd = 389	Σfd^2 =1157

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

$$= \sqrt{\frac{1157}{365} - \left(\frac{389}{365}\right)^2} \times 10$$

$$= \sqrt{3.1699 - 1.1358} \times 10$$

$$= \sqrt{2.0341} \times 10$$

$$= 1.4262 \times 10$$

$$= 14.26°c$$

Coefficient of Variation :

The coefficient of variation is obtained by dividing the standard deviation by the mean and multiply it by 100. symbolically,

$$\text{Coefficient of variation (C.V)} = \frac{\sigma}{\overline{X}} \times 100$$

Moments:

Definition of moments:

Moments can be defined as the arithmetic mean of various powers of deviations taken from the mean of a distribution. These moments are known as central moments.

The first four moments about arithmetic mean or central moments are defined below.

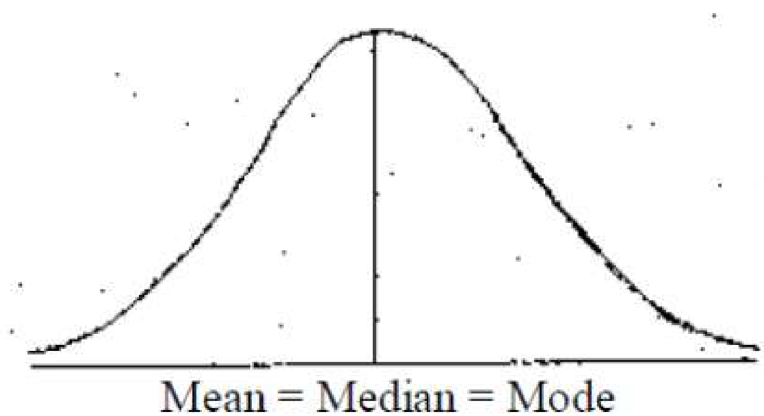
	Individual series	Discrete series
First moments about the mean; μ_1	$\frac{\Sigma(x-\overline{x})}{n} = 0$	$\frac{\Sigma f(x-\overline{x})}{N} = 0$
Second moments about the mean; μ_2	$\frac{\Sigma(x-\overline{x})^2}{n} = \sigma^2$	$\frac{\Sigma f(x-\overline{x})^2}{N}$
Third moments about the mean ; μ_3	$\frac{\Sigma(x-\overline{x})^3}{n}$	$\frac{\Sigma f(x-\overline{x})^3}{N}$
Fourth moment about the Mean ; μ_4	$\frac{\Sigma(x-\overline{x})^4}{n}$	$\frac{\Sigma f(x-\overline{x})^4}{N}$

Skewness:

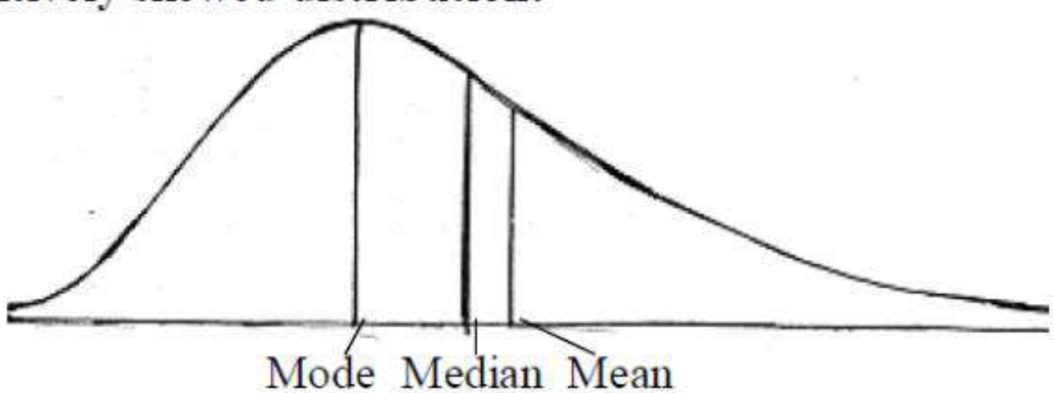
Meaning:

Skewness means ' lack of symmetry' . We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data.If in a distribution mean = median = mode, then that distribution is known as symmetrical distribution. If in a distribution mean ≠ median ≠ mode , then it is not a symmetrical distribution and it is called a skewed distribution and such a distribution could either be positively skewed or negatively skewed.

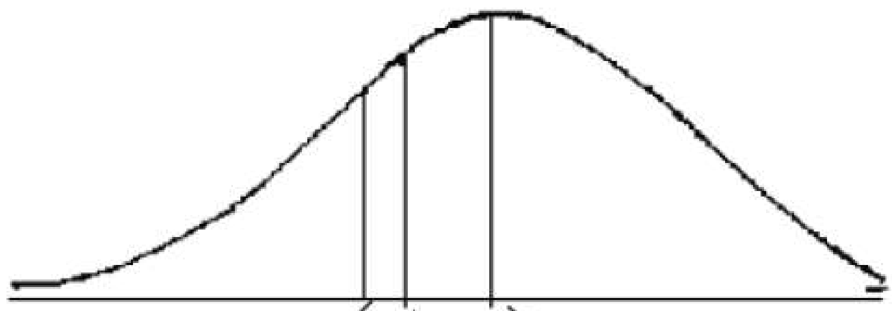
a) Symmetrical distribution:



b)Positively skewed distribution:



c) Negatively skewed distribution:



Karl – Pearson' s Coefficient of skewness:

According to Karl – Pearson, the absolute measure of skewness = mean – mode. This measure is not suitable for making valid comparison of the skewness in two or more distributions because the unit of measurement may be different in different series. To avoid this difficulty use relative measure of skewness called Karl – Pearson' s coefficient of skewness given by:

$$\text{Karl – Pearson' s Coefficient Skewness} = \frac{\text{Mean - Mode}}{S.D.}$$

Example 18:

Calculate Karl – Pearson' s coefficient of skewness for the following data.

25, 15, 23, 40, 27, 25, 23, 25, 20

Solution:

Computation of Mean and Standard deviation :

Size	Deviation from A=25 D	d^2
25	0	0
15	-10	100
23	- 2	4
40	15	225
27	2	4
25	0	0
23	-2	4
25	0	0
20	- 5	25
N=9	$\sum d=-2$	$\sum d^2=362$

$$\text{Mean} = A + \frac{\sum d}{n}$$

$$= 25 + \frac{-2}{9}$$

$$= 25 - 0.22 = 24.78$$

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

$$= \sqrt{\frac{362}{9} - \left(\frac{-2}{9}\right)^2}$$

$$= \sqrt{40.22 - 0.05}$$

$$= \sqrt{40.17} = 6.3$$

Mode = 25, as this size of item repeats 3 times
Karl – Pearson's coefficient of skewness

$$= \frac{\text{Mean - Mode}}{S.D.}$$

$$= \frac{24.78 - 25}{6.3}$$

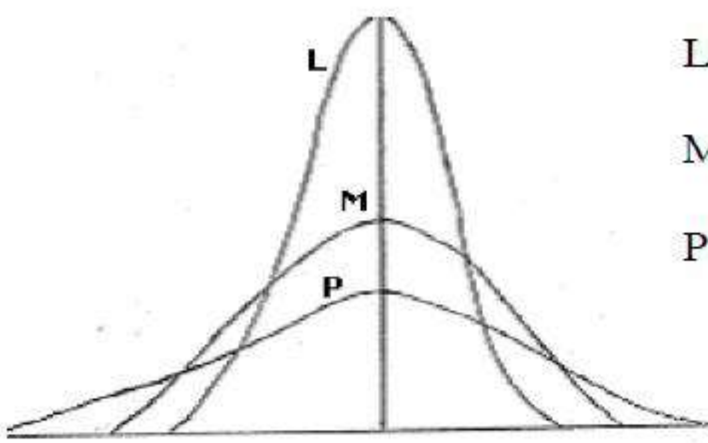
$$= \frac{-0.22}{6.3}$$

$$= -0.03$$

Kurtosis:

The expression 'Kurtosis' is used to describe the peakedness of a curve.

The three measures – central tendency, dispersion and skewness describe the characteristics of frequency distributions. But these studies will not give us a clear picture of the characteristics of a distribution.



L = Lepto Kurtic

M = Meso Kurtic

P = Platy Kurtic

Measure of Kurtosis:

The measure of kurtosis of a frequency distribution based moments is denoted by β_2 and is given by

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

If $\beta_2 = 3$, the distribution is said to be normal and the curve is mesokurtic.

If $\beta_2 > 3$, the distribution is said to be more peaked and the curve is leptokurtic.

If $\beta_2 < 3$, the distribution is said to be flat topped and the curve is platykurtic.

Example 24:

Calculate β_1 and β_2 for the following data.

X :	0	1	2	3	4	5	6	7	8
F :	5	10	15	20	25	20	15	10	5

Solution:

[**Hint:** Refer Example of page 172 and get the values of first four central moments and then proceed to find β_1 and β_2]

$\mu_1 = 0$ $\qquad$ $\mu_2 = \frac{\sum fd^2}{N} = \frac{500}{125} = 4$

$\mu_3 = \frac{\sum fd^3}{N} = 0$ $\qquad$ $\mu_4 = \frac{\sum fd^4}{N} = \frac{4700}{125} = 37.6$

$$\therefore \beta_1 = \frac{\mu^2{}_3}{\mu^3{}_2} = \frac{0}{64} = 0$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{37.6}{4^2}$$

$$= \frac{37.6}{16} = 2.35$$

The value of β_2 is less than 3, hence the curve is platykurtic.

POSSIBLE QUESTIONS(TWO MARKS)

1. Define standard deviation.

2. If the values of moments μ_2, μ_3, μ_4 are 2.83, 3.38, 30.295 then calculate the values of β_1 and β_2.

3.what is types of averages?

4. What are the merits of geometric mean?

POSSIBLE QUESTIONS(EIGHT MARKS)

1.Calculate the geometric mean for the following data:

x :	12	13	14	15	16	17
f :	5	4	4	3	2	1

2.Find the standard deviation of the following distribution:

Age :	20-25	25-30	30-35	35-40	40-45	45-50
No of persons:	170	110	80	45	40	35

3.Find the Geometric mean for the data given below:

Marks	Frequency
4-8	6
8-12	10
12-16	18
16-20	30
20-24	15

4.Calculate first four moments about the mean from the following data:

Marks :	0-10	0-20	20-30	30-40	40-50	50-60	60-70
No of students:	8	12	20	30	15	10	5

5.Calculate the value of mode using the grouping table for the following data:

Marks :	10	15	20	25	30	35	40
Frequency:	8	12	36	25	28	18	9

6.Calculate Karl Pearson's coefficient of skewness

Variable	Frequency	Variable	Frequency
70-80	11	30-40	21
60-70	22	20-30	11
50-60	30	10-20	6
40-50	35	0-10	5

7.Calculate the median from the following data

Class group	f	class group	f
100-120	6	160-170	60
120-130	25	170-180	38
130-140	48	180-190	22
140-150	72	190-200	3
150-160	116		

8.Calculate the mean deviation from the mean for the following data:

Size	2	4	6	8	10	12	14	16
Frequency	2	2	4	5	3	2	1	1

9.Calculate the median and mode of the data given below. Using them find arithmetic mean.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No of students	8	15	22	20	10	5

10.From the prices of shares of X and Y below find out which is more stable in value.

X	35	54	52	53	56	58	52	50	51	49
Y	108	107	105	105	106	107	104	103	104	101

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Post)
Coimbatore –641 021

DEPARTMENT OF MATHEMATICS

SUBJECT: STATISTICS FOR BUSINESS DECISIONS SEMESTER I L T P C

SUBJECT CODE: 17BAU102 CLASS:I UG(BBA) 5 0 0 5

UNIT II

Correlation Analysis: Meaning and significance. Correlation and Causation, Types of correlation. Methods of studying simple correlation - Scatter diagram, Karl Pearson's coefficient of correlation, Spearman's Rank correlation coefficient, Regression Analysis: Meaning and significance, Regression vs. Correlation. Linear Regression, Regression lines
(X on Y, Y on X) and Standard error of estimate.

TEXTBOOK:

1. Gupta,S.P. Statistical Method(34th e.).New Delhi: Sultan Chand & Sons

REFERENCES:

1.Navnitham .P.A .(2004).Business Mathematics and statistics.Trichy:Jai Publications.

2.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand & Company Ltd.

3 .Robert E.stine (2013) Statistics for Business: Decision Making and Analysis : Publisher: Pearson Education; 2 edition (2013)

Introduction:

The term correlation is used by a common man without knowing that he is making use of the term correlation. For example when parents advice their children to work hard so that they may get good marks, they are correlating good marks with hard work.

The study related to the characteristics of only variable such as height, weight, ages, marks, wages, etc., is known as univariate analysis. The statistical Analysis related to the study of the relationship between two variables is known as Bi-Variate Analysis. Some times the variables may be inter-related. In health sciences we study the relationship between blood pressure and age, consumption level of some nutrient and weight gain, total income and medical expenditure, etc., The nature and strength of relationship may be examined by correlation and Regression analysis.

Thus Correlation refers to the relationship of two variables or more. (e-g) relation between height of father and son, yield and rainfall, wage and price index, share and debentures etc.

Correlation is statistical Analysis which measures and analyses the degree or extent to which the two variables fluctuate with reference to each other. The word relationship is important. It indicates that there is some connection between the variables. It measures the closeness of the relationship. Correlation does not indicate cause and effect relationship. Price and supply, income and expenditure are correlated.

Definitions:

1. Correlation Analysis attempts to determine the degree of relationship between variables- Ya-Kun-Chou.
2. Correlation is an analysis of the covariation between two or more variables.- A.M.Tuttle.

Correlation expresses the inter-dependence of two sets of variables upon each other. One variable may be called as (subject) independent and the other relative variable (dependent). Relative variable is measured in terms of subject.

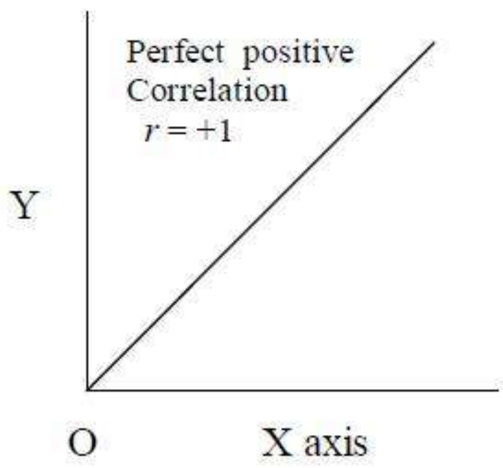
Uses of correlation:

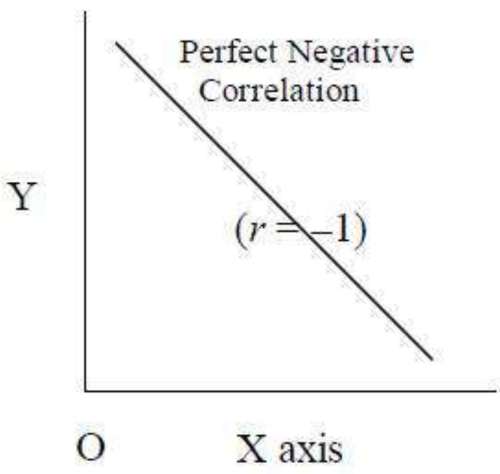
1. It is used in physical and social sciences.
2. It is useful for economists to study the relationship between variables like price, quantity etc. Businessmen estimates costs, sales, price etc. using correlation.
3. It is helpful in measuring the degree of relationship between the variables like income and expenditure, price and supply, supply and demand etc.
4. Sampling error can be calculated.
5. It is the basis for the concept of regression.

Scatter Diagram:

It is the simplest method of studying the relationship between two variables diagrammatically. One variable is represented along the horizontal axis and the second variable along the vertical axis. For each pair of observations of two variables, we put a dot in the plane. There are as many dots in the plane as the number of paired observations of two variables. The direction of dots shows the scatter or concentration of various points. This will show the type of correlation.

1. If all the plotted points form a straight line from lower left hand corner to the upper right hand corner then there is Perfect positive correlation. We denote this as $r = +1$





Linear and Non-linear correlation:

If the ratio of change between the two variables is a constant then there will be linear correlation between them.

Consider the following.

X	2	4	6	8	10	12
Y	3	6	9	12	15	18

Here the ratio of change between the two variables is the same. If we plot these points on a graph we get a straight line.

If the amount of change in one variable does not bear a constant ratio of the amount of change in the other. Then the relation is called Curvi-linear (or) non-linear correlation. The graph will be a curve.

Simple and Multiple correlation:

When we study only two variables, the relationship is simple correlation. For example, quantity of money and price level, demand and price. But in a multiple correlation we study more than two variables simultaneously. The relationship of price, demand and supply of a commodity are an example for multiple correlation.

Partial and total correlation:

The study of two variables excluding some other variable is called **Partial correlation**. For example, we study price and demand eliminating supply side. In total correlation all facts are taken into account.

Computation of correlation:

When there exists some relationship between two variables, we have to measure the degree of relationship. This measure is called the measure of correlation (or) correlation coefficient and it is denoted by ' r' .

Co-variation:

The covariation between the variables x and y is defined as

Cov(x,y) = $\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$ where $\bar{x}, \bar{y}$ are respectively means of x and y and ' n' is the number of pairs of observations.

Karl pearson' s coefficient of correlation:

Karl pearson, a great biometrician and statistician, suggested a mathematical method for measuring the magnitude of linear relationship between the two variables. It is most widely used method in practice and it is known as pearsonian coefficient of correlation. It is denoted by 'r'. The formula for calculating 'r' is

(i) $r = \dfrac{Cov(x,y)}{\sigma_x . \sigma_y}$ where σ_x, σ_y are S.D of x and y respectively.

(ii) $r = \dfrac{\Sigma xy}{n \ \sigma_x \ \sigma_y}$

(iii) $r = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$, $X = x - \bar{x}$, $Y = y - \bar{y}$

when the deviations are taken from the actual mean we can apply any one of these methods. Simple formula is the third one.

The third formula is easy to calculate, and it is not necessary to calculate the standard deviations of x and y series respectively.

Steps:

1. Find the mean of the two series x and y.
2. Take deviations of the two series from x and y.
   $X = x - \bar{x}$, $Y = y - \bar{y}$
3. Square the deviations and get the total, of the respective squares of deviations of x and y and denote by ΣX^2 , ΣY^2 respectively.
4. Multiply the deviations of x and y and get the total and Divide by n. This is covariance.
5. Substitute the values in the formula.

$$r = \frac{cov(x,y)}{\sigma x . \sigma y} = \frac{\Sigma(x-\bar{x})\ (y - \bar{y})/n}{\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} . \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}}$$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$r = \frac{9 \times 597 - 45 \times 108}{\sqrt{\left(9\times 285 - (45)^2\right).\left(9\times 1356 - (108)^2\right)}}$$

$$r = \frac{5373 - 4860}{\sqrt{(2565-2025)(12204-11664)}}$$

$$= \frac{513}{\sqrt{540 \times 540}} = \frac{513}{540} = 0.95$$

Working rule (ii) (shortcut method)

We have $r = \dfrac{Cov(x,y)}{\sigma_x.\sigma_y}$

where Cov(x,y) = $\dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$

Take the deviation from x as x – A and the deviation from y as y – B

$$\text{Cov(x,y)} = \frac{\Sigma\ [(x-A)-(\bar{x}-A)]\ [(y-B)-(\bar{y}-B)]}{n}$$

$$= \frac{1}{n}\ \Sigma\ [(x-A)\ (y-B)\ -\ (x-A)\ (\bar{y}-B)$$
$$-\ (\bar{x}-A)(y-B) + (\bar{x}-A)(\bar{y}-B)]$$

$$= \frac{1}{n}\ \Sigma\ [(x-A)\ (y-B)\ -\ (\bar{y}-B)\ \frac{\Sigma(x-A)}{n}$$
$$-\ (\bar{x}-A)\frac{\Sigma(y-B)}{n} + \frac{\Sigma(\bar{x}-A)(\bar{y}-B)}{n}$$

$$= \frac{\Sigma(x-A)(y-B)}{n} - (\bar{y}-B)\ (\bar{x} - \frac{nA}{n})$$
$$-\ (\bar{x}-A)\ (\bar{y} - \frac{nB}{n}) + (\bar{x}-A)\ (\bar{y}-B)$$

$$= \frac{\Sigma(x - A)(y - B)}{n} - (\bar{y} - B)(\bar{x} - A) - (\bar{x} - A)(\bar{y} - B) + (\bar{x} - A)(\bar{y} - B)$$

$$= \frac{\Sigma(x - A)(y - B)}{n} - (\bar{x} - A)(\bar{y} - B)$$

Let $x - A = u$; $y - B = v$; $\bar{x} - A = \bar{u}$; $\bar{y} - B = \bar{v}$

$$\therefore \text{Cov}(x,y) = \frac{\Sigma uv}{n} - \bar{u}\bar{v}$$

$$\sigma\sigma_x^2 = \frac{\Sigma u^2}{n} - \bar{u}^2 = \sigma u^2$$

$$\sigma\sigma_y^2 = \frac{\Sigma v^2}{n} - \bar{v}^2 = \sigma v^2$$

$$\therefore r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right].\left[(n\Sigma v^2) - (\Sigma v)^2\right]}}$$

Example 3:

Calculate Pearson' s Coefficient of correlation.

X	45	55	56	58	60	65	68	70	75	80	85
Y	56	50	48	60	62	64	65	70	74	82	90

X	Y	u = x-A	v = y-B	u^2	v^2	uv
45	56	-20	-14	400	196	280
55	50	-10	-20	100	400	200
56	48	-9	-22	81	484	198
58	60	-7	-10	49	100	70
60	62	-5	-8	25	64	40
65	64	0	-6	0	36	0
68	65	3	-5	9	25	-15
70	70	5	0	25	0	0
75	74	10	4	100	16	40
80	82	15	12	225	144	180
85	90	20	20	400	400	400
		2	-49	1414	1865	1393

$$r = \frac{n\Sigma uv - (\Sigma u)\ (\Sigma v)}{\sqrt{[n\Sigma u^2 - (\Sigma u^2)]\ [n\Sigma v^2 - (\Sigma v)^2]}}$$

$$r = \frac{11 \times 1393 - 2 \times (-49)}{\sqrt{(1414\times 11 - (2)^2)\times(1865\times 11 - (-49)^2)}}$$

$$= \frac{15421}{\sqrt{15550 \times 18114}} = \frac{15421}{16783.11} = +0.92$$

Limitations:

1. Correlation coefficient assumes linear relationship regardless of the assumption is correct or not.
2. Extreme items of variables are being unduly operated on correlation coefficient.
3. Existence of correlation does not necessarily indicate cause-effect relation.

Interpretation:

The following rules helps in interpreting the value of ' r' .

1. When r = 1, there is perfect +ve relationship between the variables.
2. When r = -1, there is perfect –ve relationship between the variables.
3. When r = 0, there is no relationship between the variables.
4. If the correlation is +1 or –1, it signifies that there is a high degree of correlation. (+ve or –ve) between the two variables.

If r is near to zero (ie) 0.1,-0.1, (or) 0.2 there is less correlation.

Rank Correlation:

It is studied when no assumption about the parameters of the population is made. This method is based on ranks. It is useful to study the qualitative measure of attributes like honesty, colour, beauty, intelligence, character, morality etc.The individuals in the group can be arranged in order and there on, obtaining for each individual a number showing his/her rank in the group. This method was developed by Edward Spearman in 1904. It is defined as $r = 1 - \frac{6\Sigma D^2}{n^3 - n}$ r = rank correlation coefficient.

Note: Some authors use the symbol ρ for rank correlation.

ΣD^2 = sum of squares of differences between the pairs of ranks.

n = number of pairs of observations.

The value of r lies between –1 and +1. If r = +1, there is complete agreement in order of ranks and the direction of ranks is also same. If r = -1, then there is complete disagreement in order of ranks and they are in opposite directions.

Computation for tied observations: There may be two or more items having equal values. In such case the same rank is to be given. The ranking is said to be tied. In such circumstances an average rank is to be given to each individual item. For example if the value so is repeated twice at the 5^{th} rank, the common rank to be assigned to each item is $\frac{5+6}{2} = 5.5$ which is the average of 5 and 6 given as 5.5, appeared twice.

If the ranks are tied, it is required to apply a correction factor which is $\frac{1}{12}$ (m^3-m). A slightly different formula is used when there is more than one item having the same value.

The formula is

$$r = 1 - \frac{6[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + ...]}{n^3 - n}$$

Where m is the number of items whose ranks are common and should be repeated as many times as there are tied observations.

Example 6:

In a marketing survey the price of tea and coffee in a town based on quality was found as shown below. Could you find any relation between and tea and coffee price.

Price of tea	88	90	95	70	60	75	50
Price of coffee	120	134	150	115	110	140	100

Price of tea	Rank	Price of coffee	Rank	D	D^2
88	3	120	4	1	1
90	2	134	3	1	1
95	1	150	1	0	0
70	5	115	5	0	0
60	6	110	6	0	0
75	4	140	2	2	4
50	7	100	7	0	0
					$\Sigma D^2 = 6$

$$r = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 6}{7^3 - 7}$$

$$= 1 - \frac{36}{336} = 1 - 0.1071$$

$$= 0.8929$$

The relation between price of tea and coffee is positive at 0.89. Based on quality the association between price of tea and price of coffee is highly positive.

Example 7:

In an evaluation of answer script the following marks are awarded by the examiners.

1^{st}	88	95	70	960	50	80	75	85
2^{nd}	84	90	88	55	48	85	82	72

60 is repeated 3 times in test 1.
60,65 is repeated twice in test 2.
m = 3; m = 2; m = 2

$$r = 1 - \frac{6[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)}{n^3 - n}$$

$$= 1 - \frac{6[50 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)]}{10^3 - 10}$$

$$= 1 - \frac{6[50 + 2 + 0.5 + 0.5]}{990}$$

$$= 1 - \frac{6 \times 53}{990} = \frac{672}{990} = 0.68$$

REGRESSION

9.1 Introduction:

After knowing the relationship between two variables we may be interested in estimating (predicting) the value of one variable given the value of another. The variable predicted on the basis of other variables is called the "dependent" or the ' explained' variable and the other the ' independent' or the ' predicting' variable. The prediction is based on average relationship derived statistically by regression analysis. The equation, linear or otherwise, is called the regression equation or the explaining equation.

For example, if we know that advertising and sales are correlated we may find out expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales.

The relationship between two variables can be considered between, say, rainfall and agricultural production, price of an input and the overall cost of product, consumer expenditure and disposable income. Thus, regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

9.1.1 Definition:

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.

9.2 Types Of Regression:

The regression analysis can be classified into:

a) Simple and Multiple
b) Linear and Non –Linear
c) Total and Partial

a) Simple and Multiple:

In case of simple relationship only two variables are considered, for example, the influence of advertising expenditure on sales turnover. In the case of multiple relationship, more than two variables are involved. On this while one variable is a dependent variable the remaining variables are independent ones.

For example, the turnover (y) may depend on advertising expenditure (x) and the income of the people (z). Then the functional relationship can be expressed as $y = f(x,z)$.

b) Linear and Non-linear:

The linear relationships are based on straight-line trend, the equation of which has no-power higher than one. But, remember a linear relationship can be both simple and multiple. Normally a linear relationship is taken into account because besides its simplicity, it has a better predective value, a linear trend can be easily projected into the future. In the case of non-linear relationship curved trend lines are derived. The equations of these are parabolic.

c) Total and Partial:

In the case of total relationships all the important variables are considered. Normally, they take the form of a multiple relationships because most economic and business phenomena are affected by multiplicity of cases. In the case of partial relationship one or more variables are considered, but not all, thus excluding the influence of those not found relevant for a given purpose.

9.3 Linear Regression Equation:

If two variables have linear relationship then as the independent variable (X) changes, the dependent variable (Y) also changes. If the different values of X and Y are plotted, then the two straight lines of best fit can be made to pass through the plotted points. These two lines are known as regression lines. Again, these regression lines are based on two equations known as regression equations. These equations show best estimate of one variable for the known value of the other. The equations are linear.

Linear regression equation of Y on X is

$$Y = a + bX \;\ldots\ldots(1)$$

And X on Y is

$$X = a + bY \;\ldots\ldots(2)$$

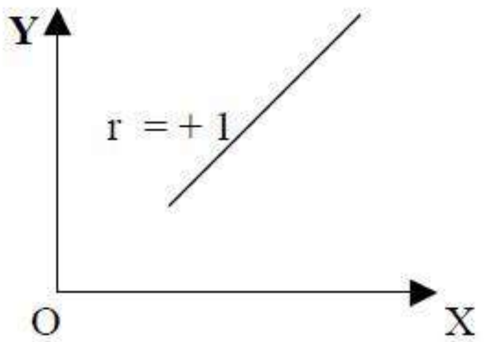
a, b are constants.

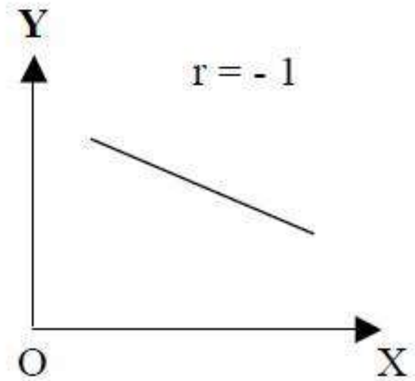
From (1) We can estimate Y for known value of X.
(2) We can estimate X for known value of Y.

9.3.1 Regression Lines:

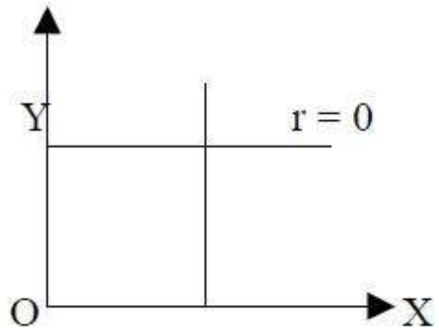
For regression analysis of two variables there are two regression lines, namely Y on X and X on Y. The two regression lines show the average relationship between the two variables.

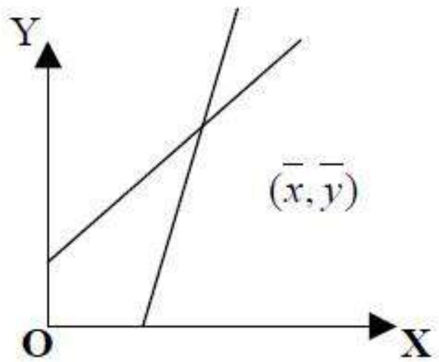
For perfect correlation, positive or negative i.e., $r = \pm 1$, the two lines coincide i.e., we will find only one straight line. If r = 0, i.e., both the variables are independent then the two lines will cut each other at right angle. In this case the two lines will be parallel to X and Y-axes.





Lastly the two lines intersect at the point of means of X and Y. From this point of intersection, if a straight line is drawn on X-axis, it will touch at the mean value of x. Similarly, a perpendicular drawn from the point of intersection of two regression lines on Y-axis will touch the mean value of Y.





9.3.2 Principle of ' Least Squares' :

Regression shows an average relationship between two variables, which is expressed by a line of regression drawn by the method of "least squares". This line of regression can be derived graphically or algebraically. Before we discuss the various methods let us understand the meaning of least squares.

A line fitted by the method of least squares is known as the line of best fit. The line adapts to the following rules:

(i) The algebraic sum of deviation in the individual observations with reference to the regression line may be equal to zero. i.e.,

$\Sigma(X - Xc) = 0$ or $\Sigma(Y - Yc) = 0$

Where Xc and Yc are the values obtained by regression analysis.

(ii) The sum of the squares of these deviations is less than the sum of squares of deviations from any other line. i.e.,

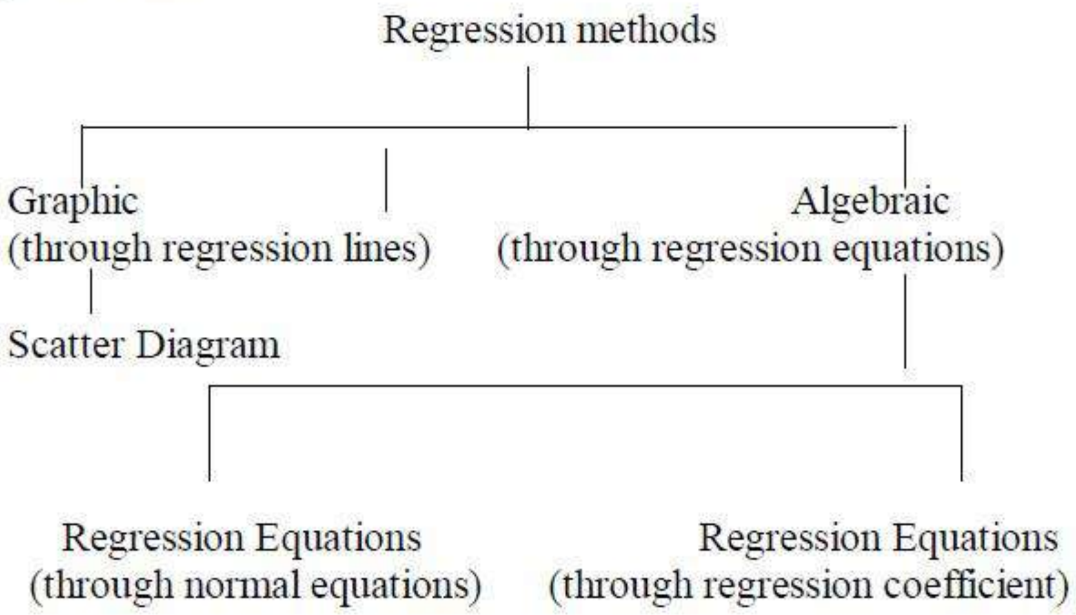
$\Sigma(Y - Yc)^2 < \Sigma(Y - Ai)^2$

Where Ai = corresponding values of any other straight line.

(iii) The lines of regression (best fit) intersect at the mean values of the variables X and Y, i.e., intersecting point is $\bar{x}, \bar{y}$.

9.4 Methods of Regression Analysis:

The various methods can be represented in the form of chart given below:



9.4.1 Graphic Method:

Scatter Diagram:

Under this method the points are plotted on a graph paper representing various parts of values of the concerned variables. These points give a picture of a scatter diagram with several points spread over. A regression line may be drawn in between these points either by free hand or by a scale rule in such a way that the squares of the vertical or the horizontal distances (as the case may be) between the points and the line of regression so drawn is the least. In other words, it should be drawn faithfully as the line of best fit leaving equal number of points on both sides in such a manner that the sum of the squares of the distances is the best.

9.4.2 Algebraic Methods:

(i) Regression Equation.

The two regression equations

for X on Y; $X = a + bY$

And for Y on X; $Y = a + bX$

Where X, Y are variables, and a,b are constants whose values are to be determined

For the equation, $X = a + bY$

The normal equations are

$$\sum X = na + b\sum Y \text{ and}$$

$$\sum XY = a\sum Y + b\sum Y^2$$

For the equation, $Y = a + bX$, the normal equations are

$$\sum Y = na + b\sum X \text{ and}$$

$$\sum XY = a\sum X + b\sum X^2$$

From these normal equations the values of a and b can be determined.

Example 1:

Find the two regression equations from the following data:

X:	6	2	10	4	8
Y:	9	11	5	8	7

Solution:

X	Y	X^2	Y^2	XY
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
30	40	220	340	214

Regression equation of Y on X is $Y = a + bX$ and the normal equations are

$$\sum Y = na + b\sum X$$
$$\sum XY = a\sum X + b\sum X^2$$

Substituting the values, we get

$$40 = 5a + 30b \text{ ..... } (1)$$
$$214 = 30a + 220b \text{ ......}(2)$$

Multiplying (1) by 6

$$240 = 30a + 180b \text{......}(3)$$

(2) – (3) $\quad -26 = 40b$

or $b = -\dfrac{26}{40} = -0.65$

Now, substituting the value of 'b' in equation (1)

$40 = 5a - 19.5$

$5a = 59.5$

$a = \dfrac{59.5}{5} = 11.9$

Hence, required regression line Y on X is $Y = 11.9 - 0.65X$.

Again, regression equation of X on Y is

$X = a + bY$ and

The normal equations are

$$\sum X = na + b\sum Y \text{ and}$$
$$\sum XY = a\sum Y + b\sum Y^2$$

Now, substituting the corresponding values from the above table, we get

$$30 = 5a + 40b \ldots(3)$$
$$214 = 40a + 340b \ldots(4)$$

Multiplying (3) by 8, we get

$$240 = 40a + 320\,b \ldots(5)$$

(4) – (5) gives

$$-26 = 20b$$
$$b = -\frac{26}{20} = -1.3$$

Substituting $b = -1.3$ in equation (3) gives

$$30 = 5a - 52$$
$$5a = 82$$
$$a = \frac{82}{5} = 16.4$$

Hence, Required regression line of X on Y is

X = 16.4 – 1.3Y

(ii) Regression Co-efficents:

The regression equation of Y on X is $y_e = \overline{y} + r\frac{\sigma_y}{\sigma_x}(x - \overline{x})$

Here, the regression Co.efficient of Y on X is

$$b_1 = b_{yx} = r\frac{\sigma_y}{\sigma_x}$$
$$y_e = \overline{y} + b_1(x - \overline{x})$$

The regression equation of X on Y is

$$X_e = \overline{x} + r\frac{\sigma_x}{\sigma_y}(y - \overline{y})$$

Here, the regression Co-efficient of X on Y

$$b_2 = b_{xy} = r\frac{\sigma_x}{\sigma_y}$$
$$X_e = \overline{X} + b_2(y - \overline{y})$$

If the deviation are taken from respective means of x and y

$$b_1 = b_{yx} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (X-\overline{X})^2} = \frac{\sum xy}{\sum x^2} \quad \text{and}$$

$$b_2 = b_{xy} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (Y-\overline{Y})^2} = \frac{\sum xy}{\sum y^2}$$

where $x = X - \overline{X}, y = Y - \overline{Y}$

If the deviations are taken from any arbitrary values of x and y (short – cut method)

$$b_1 = b_{yx} = \frac{n\sum uv - \sum u \sum v}{n\sum u^2 - \left(\sum u\right)^2}$$

$$b_2 = b_{xy} = \frac{n\sum uv - \sum u \sum v}{n\sum v^2 - \left(\sum v\right)^2}$$

where u = x – A : v = Y-B

A = any value in X

B = any value in Y

9.5 Properties of Regression Co-efficient:

1. Both regression coefficients must have the same sign, ie either they will be positive or negative.
2. correlation coefficient is the geometric mean of the regression coefficients ie, $r = \pm\sqrt{b_1 b_2}$
3. The correlation coefficient will have the same sign as that of the regression coefficients.
4. If one regression coefficient is greater than unity, then other regression coefficient must be less than unity.
5. Regression coefficients are independent of origin but not of scale.
6. Arithmetic mean of b_1 and b_2 is equal to or greater than the coefficient of correlation. Symbolically $\frac{b_1 + b_2}{2} \geq r$

7. If r=0, the variables are uncorrelated , the lines of regression become perpendicular to each other.
8. If r= ± 1, the two lines of regression either coincide or parallel to each other
9. Angle between the two regression lines is $\theta = \tan^{-1}\left[\frac{m_1 - m_2}{1 + m_1 m_2}\right]$

   where m_1 and,m_2 are the slopes of the regression lines X on Y and Y on X respectively.
10. The angle between the regression lines indicates the degree of dependence between the variables.

Example 2:

If 2 regression coefficients are $b_1=\frac{4}{5}$ and $b_2 =\frac{9}{20}$ .What would be the value of r?

Solution:

The correlation coefficient , $r = \pm\sqrt{b_1 b_2}$

$$= \sqrt{\frac{4}{5} \text{x} \frac{9}{20}}$$

$$= \sqrt{\frac{36}{100}} = \frac{6}{10} = 0.6$$

Example 3:

Given $b_1 = \frac{15}{8}$ and $b_2 = \frac{3}{5}$, Find r

Solution:

$r = \pm\sqrt{b_1 b_2}$

$$= \sqrt{\frac{15}{8} \text{x} \frac{3}{5}}$$

$$= \sqrt{\frac{9}{8}} \quad =1.06$$

It is not possible since r, cannot be greater than one. So the given values are wrong

Example 6:

In a correlation study, the following values are obtained

	X	Y
Mean	65	67
S.D	2.5	3.5

Co-efficient of correlation = 0.8
Find the two regression equations that are associated with the above values.

Solution:

Given,

$\overline{X} = 65$, $\overline{Y} = 67$, $\sigma_x = 2.5$, $\sigma_y = 3.5$, $r = 0.8$

The regression co-efficient of Y on X is

$$b_{yx} = b_1 = r\frac{\sigma_y}{\sigma_x}$$

$$= 0.8 \times \frac{3.5}{2.5} = 1.12$$

The regression coefficient of X on Y is

$$b_{xy} = b_2 = r\frac{\sigma_x}{\sigma_y}$$

$$= 0.8 \times \frac{2.5}{3.5} = 0.57$$

Hence, the regression equation of Y on X is

$$Y_e = \overline{Y} + b_1(X - \overline{X})$$

$= 67 + 1.12\,(X-65)$

$= 67 + 1.12\,X - 72.8$

$= 1.12X - 5.8$

The regression equation of X on Y is

$$X_e = \overline{X} + b_2(Y - \overline{Y})$$

$= 65 + 0.57\,(Y-67)$

$= 65 + 0.57Y - 38.19$

$= 26.81 + 0.57Y$

POSSIBLE QUESTIONS (TWO MARKS)

1. What are the features of Spearman's correlation coefficient?
2. What do you mean by regression equations?
3. What are the methods of studying correlation?
4. What are the properties of linear regression?

POSSIBLE QUESTIONS(EIGHT MARKS)

1.Calculate Karl Pearson's correlation coefficient between the marks in English and Hindi obtained by 10 students:

Marks in English :	10	25	13	25	22	11	12	25	21	20
Marks in Hndi:	12	22	16	15	18	18	17	23	24	17

2.From the following data calculate the regression equations taking deviation of items from the mean of X and Y series:

X:	6	2	10	4	8
Y:	9	11	5	8	7

3.Calculate Spearman's coefficient of correlation between mark assigned to ten students by judges X and Y in a certain competitive test as shown below:

S. No :	1	2	3	4	5	6	7	8	9	10
Marks by Judge X :	52	53	42	60	45	41	37	38	25	27
Marks by Judge Y :	65	68	43	38	77	48	35	30	25	50

4.The following data, based on 450 students, are given for marks in statistics and economics at a certain examination:

Mean marks in Statistics	40
Mean marks in Economics	48
S.D of marks in Statistics	12
The variance of marks in Economics	256
Some of the products of deviation of marks from their respective mean	42075

Give the equations of the two lines of regression and estimate the average marks in Economics of candidate who obtained 50 marks in statistics.

5.Explain the types of correlation.

6.Given the following data: $\overline{X}$=36, $\overline{Y}$=85, $\sigma_x = 11, \sigma_y = 8, r = 0.66$

Find the two regression equations and estimate the value of X when Y=75

7.Distinguish between Regression and Correlation.

8.The following table gives the age of cars of a certain make and actual maintenance costs. Obtain the regression equation for costs related to age.

Age of car(years)	2	4	6	8
Maintenance cost (Rs. hundred)	10	20	25	30

9.For the following data calculate the rank correlation coefficient between X and Y.

X :	1	2	3	4	5	6	7	8	9	10	11	12
Y :	12	9	6	10	3	5	4	7	8	2	11	1

10.Find two regression equations for the following two series, what is most likely value of Xwhen Y = 20 and most likely value of Y when X = 22.

X:	35	25	29	31	27	24	33	36
Y:	23	27	26	21	24	20	29	30

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Post)
Coimbatore –641 021

DEPARTMENT OF MATHEMATICS

SUBJECT: STATISTICS FOR BUSINESS DECISIONS SEMESTER: I L T P C
SUBJECT CODE: 17BAU102 CLASS:I UG (BBA) 5 0 0 5

UNIT III
Analysis of Time Series: Meaning and significance. Utility, Components of time series, Models (Additive and Multiplicative), Measurement of trend: Method of least squares, Parabolic trend and logarithmic trend.

TEXTBOOK:

1. Gupta,S.P. Statistical Method(34^{th} e.).New Delhi: Sultan Chand & Sons

REFERENCES:

1.Dr.Arora ,P.N.(1997).A foundation course statistics . New Delhi:S.Chand &Company Ltd.

2 .Robert E.stine (2013) Statistics for Business: Decision Making and Analysis :
Publisher: Pearson Education; 2 edition (2013)

Introduction:

Arrangement of statistical data in chronological order ie., in accordance with occurrence of time, is known as “Time Series”. Such series have a unique important place in the field of Economic and Business statistics. An economist is interested in estimating the likely population in the coming year so that proper planning can be carried out with regard to food supply, job for the people etc. Similarly, a business man is interested in finding out his likely sales in the near future, so that the businessman could adjust his production accordingly and avoid the possibility of inadequate production to meet the demand. In this connection one usually deal with statistical data, which are collected, observed or recorded at successive intervals of time. Such data are generally referred to as ‘ time series’ .

Definition:

According to Mooris Hamburg “A time series is a set of statistical observations arranged in chronological order”

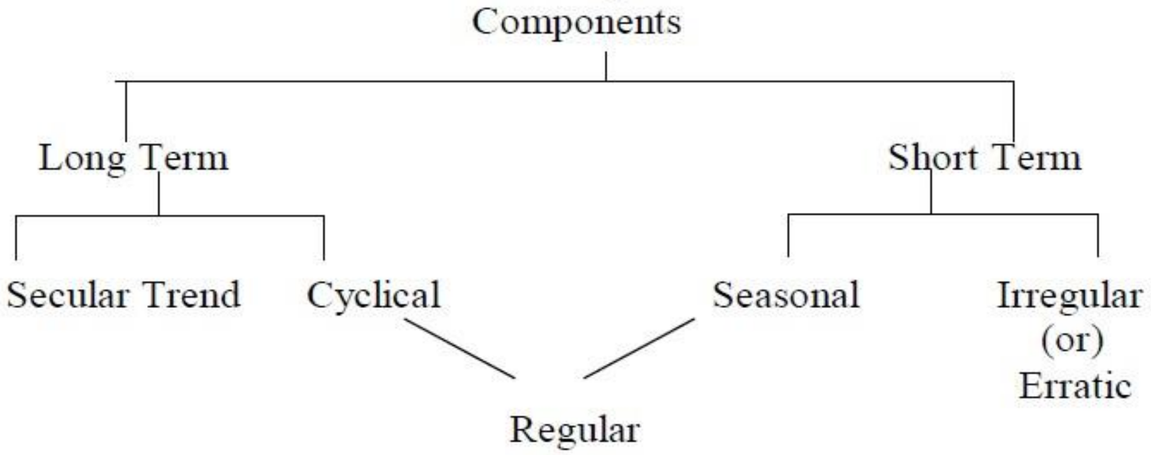
Ya-Lun- chou defining the time series as “A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables. A time series is a set of observations of a variable usually at equal intervals of time. Here time may be yearly, monthly, weekly, daily or even hourly usually at equal intervals of time.

Hourly temperature reading, daily sales, monthly production are examples of time series. Number of factors affect the observations of time series continuously, some with equal intervals of time and others are erratic studying, interpreting analyzing the factors is called Analysis of Time Series.

The Primary purpose of the analysis of time series is to discover and measure all types of variations which characterise a time series. The central objective is to decompose the various elements present in a time series and to use them in business decision making.

Components of Time series:

The components of a time series are the various elements which can be segregated from the observed data. The following are the broad classification of these components.



In time series analysis, it is assumed that there is a multiplicative relationship between these four components. Symbolically,

$$Y = T \times S \times C \times I$$

Where Y denotes the result of the four elements; T = Trend ; S = Seasonal component; C = Cyclical components; I = Irregular component

In the multiplicative model it is assumed that the four components are due to different causes but they are not necessarily independent and they can affect one another.

Another approach is to treat each observation of a time series as the sum of these four components. Symbolically

$$Y = T + S + C + I$$

The additive model assumes that all the components of the time series are independent of one another.

1) Secular Trend or Long - Term movement or simply Trend
2) Seasonal Variation
3) Cyclical Variations
4) Irregular or erratic or random movements(fluctuations)

1. **Secular Trends:** Secular trends is also called long – term trend or trend, simply. The overall nature of the series is the trend. The general tendency of a series is to increase or decrease over a period of time. Increasing trend is observed in population, price, production, literacy, etc. There is

decreasing trend in birth rate, death rate, poverty, illiteracy, etc. It is very rare to find a time series which neither increases nor decreases.

Mathematically, trend may be

(i) Linear or

(ii) Non – linear.

Graphically, linear trend is a straight line. The discussion in this chapter is restricted to linear trend. Parabolic trend equation, if necessary, can be formed as explained in 'Method of Least Squares'.

2. **Seasonal Fluctuations.** Season is a period which is less than one year. It may be a period of 6 months or 4 months or 3 months or 1 months, etc. Certain nature is observed in the first season, another nature is observed in a season in every year. In other words, the different natures recur year after year at the respective seasons. These variation over time are called seasonal fluctuations.

The factor which cause seasonal variations are of the following two kinds:

(i) Climate and weather conditions.

(ii) Customs, traditions and habits of the people.

(iii) Climate and weather condition: Sales of ice – cream, khadi and cotton clothes, etc. are more during summer. Sales of umberellas are at its peak during rainy season. Production of paddy, wheat, etc. is more in a few months and less in other months of a year. Climate and weather cause this kind of variations.

iv) Customs, traditions and habits of the people. Sales of crackers and fire works is found to be more during Deepavali every year. Cloth shops register very good sales during festival; seasons such as Deepavali, Pongal, Ramzan and Chritmas and marriage seasons. Post men are very busy in those days in sorting and delivering greeting. All these variations in sales, work load, etc. are due to the customs, traditions and habits of the people.

fluctuations a nature of the series recurs at an interval of one year. Cyclical fluctuations recur at an interval of 3 or more years. The fitting example is business cycle. In Economics and Business, there are many times series which have certain wave – like movements called business cycles, in one period, profits are easily made and are made in plenty also. Prices are high. This period is called prosperity. After this (peak) conditions things decline instead of improving. High wages, decreasing efficiently, increasing interest rate, etc. cause the decline. This is the period of recession. After touching the bottom which is called depression the condition improves. The recovery from depression leads to prosperity. The four phase of a business cycle, namely, (i) prosperity (ii) recession (iii) depression and (iv) recovery recur one after another regularly.

3. **Irregular Variations.** Variations which no not come under the other three components are called irregular variations. The other three components have certain regularity. But this is irregular. Fire, floods, earthquakes, wars, lock – outs, strikes, etc, cause irregular variations. Sometimes causes as above for irregular variations are known. Sometimes causes may not be known. For example, there may be very poor sales on a particular day in a leading cloth shop on the eve of Deepavali. Cause for such a happening may not be known. Irregular variations is called **random variation or erratic fluctuation.**

Secular Trend:

It is a long term movement in Time series. The general tendency of the time series is to increase or decrease or stagnate during a long period of time is called the secular trend or simply trend. Population growth, improved technological progress, changes in consumers taste are the various factors of upward trend. We may notice downward trend relating to deaths, epidemics, due to improved medical facilities and sanitations. Thus a time series shows fluctuations in the upward or downward direction in the long run.

Methods of Measuring Trend:

Trend is measured by the following mathematical methods.

1. Graphical method
2. Method of Semi-averages
3. Method of moving averages
4. Method of Least Squares

Method of Least Square:

This method is widely used. It plays an important role in finding the trend values of economic and business time series. It helps for forecasting and predicting the future values. The trend line by this method is called the line of best fit.

The equation of the trend line is $y = a + bx$, where the constants a and b are to be estimated so as to minimize the sum of the squares of the difference between the given values of y and the estimate values of y *by* using the equation. The constants can be obtained by solving two normal equations.

$$\Sigma y = na + b\Sigma x \quad \ldots\ldots\ldots (1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \ldots\ldots\ldots (2)$$

Here x represent time point and y are observed values. ‘n’ is the number of pair- values.

When odd number of years are given

Step 1: Writing given years in column 1 and the corresponding sales or production etc in column 2.

Step 2: Write in column 3 start with 0, 1, 2 .. against column 1 and denote it as X

Step 3: Take the middle value of X as A

Step 4: Find the deviations $u = X - A$ and write in column 4

Step 5: Find u^2 values and write in column 5.

Step 6: Column 6 gives the product uy

Now the normal equations become

$$\Sigma y = na + b\Sigma u \quad (1) \qquad \text{where } u = X\text{-}A$$

$$\Sigma uy = a\Sigma u + b\Sigma u^2 \quad (2)$$

Since $\Sigma u = 0$, From equation (1)

$$a = \frac{\Sigma y}{n}$$

From equation (2)

$$\Sigma uy = b\Sigma u^2$$

$$\therefore b = \frac{\Sigma uy}{\Sigma u^2}$$

$\therefore$ The fitted straight line is

$$y = a + bu \quad = a + b\,(X - A)$$

Example 6:

For the following data, find the trend values by using the method of Least squares

Year	1990	1991	1992	1993	1994
Production (in tones)	50	55	45	52	54

Estimate the production for the year 1996

Solution:

Year (x)	Production (y)	X= x -1990	u = X-A = X-2	u^2	uy	Trend values
1990	50	0	-2	4	-100	50.2
1991	55	1	-1	1	-55	50.7
1992	45	**2 A**	0	0	0	51.2
1993	52	3	1	1	52	51.7
1994	54	4	2	4	108	52.2
Total	256			10	5	

Where A is an assumed value

The equation of straight line is

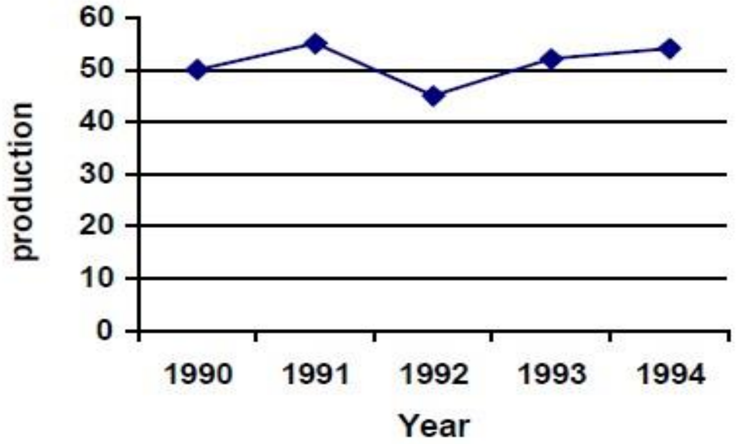
$$Y = a + bX$$

$$= a + bu, \quad \text{where } u = X - 2$$

the normal equations are

$$\Sigma y = na + b\Sigma u \text{.........}(1)$$
$$\Sigma uy = a\Sigma u + b\Sigma u^2 \text{ ....}(2)$$
since $\Sigma u = 0$ from(1) $\Sigma y = na$
$$a = \frac{\Sigma y}{n} = \frac{256}{5} = 51.2$$
From equation (2)
$$\Sigma uy = b\Sigma u^2$$
$$5 = 10b$$
$$b = \frac{5}{10} = 0.5$$
The fitted straight line is
$$y = a + bu$$
$$y = 51.2 + 0.5\ (X-2)$$
$$y = 51.2 + 0.5X - 1.0$$
$$y = 50.2 + 0.5X$$
Trend values are, 50.2, 50.7, 51.2, 51.7, 52.2

The estimate production in 1996 is put $X = x - 1990$
$$X = 1996 - 1990 = 6$$
$$Y = 50.2 + 0.5X = 50.2 + 0.5(6)$$
$$= 50.2 + 3.0 = 53.2 \text{ tonnes.}$$



When **even number of years** are given

Here we take the mean of middle two values of X as A

Then $u = \dfrac{X - A}{1/2} = 2\ (X\text{-}A)$. The other steps are as given in the odd number of years.

Example 7:

Fit a straight line trend by the method of least squares for the following data.

Year	1983	1984	1985	1986	1987	1988
Sales (Rs. in lakhs)	3	8	7	9	11	14

Also estimate the sales for the year 1991

Solution:

Year (x)	Sales (y)	X = x-1983	u =2X-5	u2	uy	Trend values
1983	3	0	-5	25	-15	3.97
1984	8	1	-3	9	-24	5.85
1985	7	2	-1	1	-7	7.73
1986	9	3	1	1	9	9.61
1987	11	4	3	9	33	11.49
1988	14	5	5	25	70	13.37
Total	52		0	70	66	

$$u = \frac{X - A}{1/2}$$

$$= 2\,(X - 2.5) = 2X - 5$$

The straight line equation is

$$y = a + bX \quad = a + bu$$

The normal equations are

$$\Sigma y = na \ \ldots\ldots(1)$$

$$\Sigma uy = b\Sigma u^2 \ \ldots\ldots(2)$$

From (1) $52 = 6a$

$$a = \frac{52}{6}$$
$$= 8.67$$

From (2) 66 = 70 b

$$b = \frac{66}{70}$$
$$= 0.94$$

The fitted straight line equation is

$$y = a+bu$$
$$y = 8.67+0.94(2X-5)$$
$$y = 8.67 + 1.88X - 4.7$$
$$y = 3.97 + 1.88X \text{ -----------(3)}$$

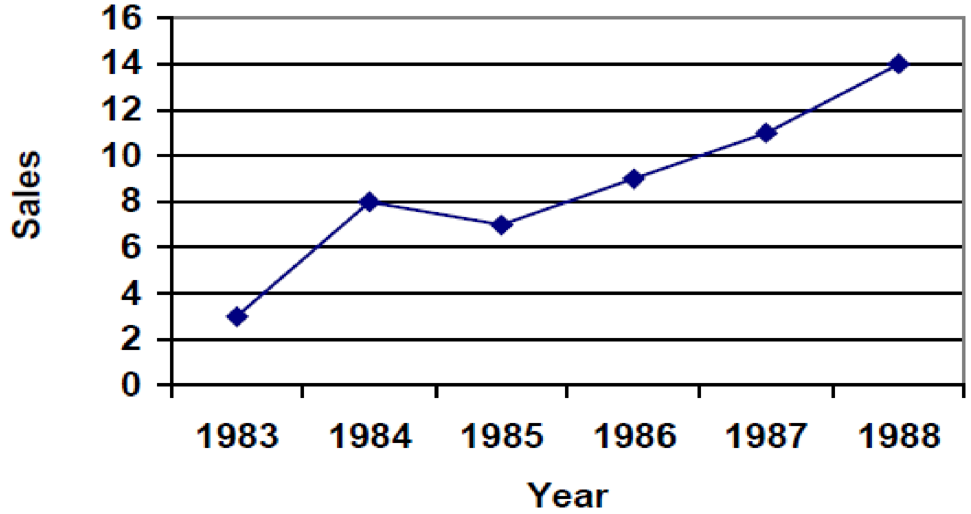
The trend values are

Put $X = 0$, $y = 3.97$ $X = 1$, $y = 5.85$
$X = 2$, $y = 7.73$ $X = 3$, $y = 9.61$
$X = 4$, $y = 11.49$ $X = 5$, $y = 13.37$

The estimated sale for the year 1991 is; put X = x –1983

$$= 1991 - 1983 = 8$$
$$y = 3.97 + 1.88 \times 8$$
$$= 19.01 \text{ lakhs}$$

The following graph will show clearly the trend line.



Merits:

1. Since it is a mathematical method, it is not subjective so it eliminates personal bias of the investigator.
2. By this method we can estimate the future values. As well as intermediate values of the time series.
3. By this method we can find all the trend values.

Demerits:

1. It is a difficult method. Addition of new observations makes re-calculations.
2. Assumption of straight line may sometimes be misleading since economics and business time series are not linear.
3. It ignores cyclical, seasonal and irregular fluctuations.
4. The trend can estimate only for immediate future and not for distant future.

Parabolic Trend Model

The curvilinear relationship for estimating the value of a dependent variable
y from an independent variable *x* might take the form

$\hat{y} = a + bx + cx^2$.This trend line is called the *parabola*.

For a non-linear equation $\hat{y} = a + bx + cx^2$, the values of constants *a*, *b*, and *c* can be determined by solving three normal equations.

$\Sigma y = na + b\Sigma x + c\Sigma x2$
$\Sigma xy = a\Sigma x + b\Sigma x2 + c\Sigma x3$
$\Sigma x2y = a\Sigma x2 + b\Sigma x3 + c\Sigma x4$

When the data can be coded so that $\Sigma x = 0$ and $\Sigma x3 = 0$, two term in the above expressions drop out and we have

$\Sigma y = na + c\Sigma x2$
$\Sigma xy = b\Sigma x2$
$\Sigma x2y = a\Sigma x2 + c\Sigma x4$

To find the exact estimated value of the variable *y*, the values of constants *a*, *b*, and *c* need to be calculated. The values of these constants can be calculated by using the following shortest method:

$$a = \frac{\sum y - c\sum x^2}{n}; b = \frac{\sum xy}{\sum x^2} \, and \, c = \frac{n\sum x^2 y - \sum x^2 \sum y}{n\sum x^4 - (\sum x^2)^2}$$

Example : The prices of a commodity during 1999-2004 are given below. Fit a parabola to these data. Estimate
the price of the commodity for the year 2005.

Year	Price	Year	Price
1999	100	2002	140
2000	107	2003	181
2001	128	2004	192

Solution: To fit a parabola *y*ˆ = a + *bx* + *cx*2, the calculations to determine the values of constants *a*, *b*, and *c* are shown in table

Calculations for Parabola Trend Line

Year	Time Scale (x)	Price (y)	x^2	x^3	x^4	xy	x^2y	Trend Values (*y*ˆ)
1999	-2	100	4	-8	16	-200	400	97.72
2000	-1	107	1	-1	1	-107	107	110.34
2001	0	128	0	0	0	0	0	126.68
2002	1	140	1	1	1	140	140	146.50
2003	2	181	4	8	16	362	724	169.88
2004	3	192	9	27	81	576	1728	196.82
	3	**848**	**19**	**27**	**115**	**771**	**3099**	**847.94**

(i) Σy = na- bΣx + cΣx2

848 = 6*a* + 3*b* + 19*c*

(ii) Σxy = aΣx + bΣx2 + cΣx3

$771 = 3a + 19b + 27c$

(iii) $\Sigma x2y = a\Sigma x2 + b\Sigma x2 + c\Sigma x4$

$3099 = 19a + 27b + 115c$

Eliminating a from eqns. (i) and (ii), we get

(iv) $694 = 35b + 35c$

Eliminating a from eqns. (ii) and (iii), we get

(v) $5352 = 280b + 168c$

Solving eqns. (iv) and (v) for b and c we get b =18.04 and c = 1.78.

Substituting values of b and c in eqn. (i), we get a = 126.68.

Hence, the required non-linear trend line becomes

$y = 126.68 + 18.04x + 1.78x^2$

Exponential Trend Model

When the given values of dependent variable y from approximately a geometric progression while the corresponding independent variable x values form an arithmetic progression, the relationship between variables x and y is given by an exponential function, and the best fitting curve is said to describe the *exponential trend.* Data from the fields of biology, banking, and economics frequently exhibit such a trend. For example, growth of bacteria, money accumulating at compound interest, sales or earnings over a short period, and so on, follow exponential growth.

The characteristics property of this law is that the rate of growth, that is, the rate of change of y with respect to x is proportional to the values of the function. The following function has this property.

y = abcx, a > 0

The letter b is a fixed constant, usually either 10 or e, where a is a constant to be determined from the data.

To assume that the law of growth will continue is usually unwarranted, so only short range predictions can be made with any considerable degree or reliability.

If we take logarithms (with base 10) of both sides of the above equation, we obtain

$\text{Log } y = \log a + (c \log b)\, x$ (7.2)

For b =10, log b =1, but for b=e, log b =0.4343 (approx.). In either case, this equation is of the form $y' = c + dx$

Where $y' = \log y$, $c = \log a$, and $d = c \log b$.

Equation (7.2) represents a straight line. A method of fitting an exponential trend line to a set of observed values of y is to fit a straight trend line to the logarithms of the y-values.

In order to find out the values of constants a and b in the exponential function, the two normal equations to be solved are

$\Sigma \log y = n \log a + \log b \Sigma x$

$\Sigma x \log y = \log a \Sigma x + \log b \Sigma x2$

When the data is coded so that $\Sigma x = 0$, the two normal equations become

$\Sigma \log y = n \log a$ or $\log a = \frac{1}{n} \Sigma \log y$

and $\Sigma x \log y = \log b \, \Sigma x^2$ or $\log b = \frac{\Sigma x \log y}{\Sigma x2}$

Coding is easily done with time-series data by simply designating the center of the time period as $x = 0$, and have equal number of plus and minus period on each side which sum to zero.

Example :

The sales (Rs. In million) of a company for the years 1995 to 1999 are:

Year :	1995	1996	1997	1998	1999
Sales :	1.6	4.5	13.8	40.2	125.0

Find the exponential trend for the given data and estimate the sales for 2002.

Solution:

computational time can be reduced by coding the data. For this consider $u = x$-3. The necessary computations are shown in table

Fitting the Exponential Trend Line

Year	Time Period x	u=x-3	u^2	Sales y	Log y	u log y
1995	1	-2	4	1.60	0.2041	-0.4082
1996	2	-1	1	4.50	0.6532	-0.6532
1997	3	0	0	13.80	1.1390	0

1998	4	1	1	40.20	1.6042	1.6042
1999	5	2	4	125.00	2.0969	4.1938
			10		**5.6983**	**4.7366**

log a $\frac{1}{n}\Sigma$ log y = $\frac{1}{5}$(5.6983) = 1.1397

Therefore log y = log a + (x+3) log b = 1.1397 + 0.4737x

For sales during 2002, x =3, and we obtain

log y = 1.1397 + 0.4737 (3) = 2.5608

y = antilog (2.5608) = 363.80

POSSIBLE QUESTIONS (TWO MARKS)

1. Define time series.
2. What are the various methods used in determining trend?
3. What are the limitations of method of least squares?
4. What are the merits of method of least squares?

POSSIBLE QUESTIONS (EIGHT MARKS)

1.Explain the components of time series.

2.The sales of a company in lakhs of rupee for the years 1990 to 1996 are given below:

Year :	1990	1991	1992	1993	1994	1995	1996
Sales:	32	47	65	88	132	190	275

Find trend values by using the equation $Y_c = ab^X$ and estimate the value for 1997

3.Fit a straight line trend by the method of least square to the following data. Assuming that the same rate of change continues what would be the predicted earnings for the year 1989.

Year :	1981	1982	1983	1984	1985	1986	1987	1988
Earnings: (Rs. Lakhs)	38	40	65	72	69	60	87	95

4.The following table give the profits of a concern for 5 years ending 1996:

Year :	1992	1993	1994	1995	1996
Profits :	1.6	4.5	13.8	40.2	125.0

(in Rs. thousands)

Fit an equation of the type $Y_c = ab^X$.

5.Fit a second degree parabola to the following data and also estimate the value for 1990 and give your comments:

Year :	1955	1960	1965	1970	1975	1980	1985
Production('000 units):	6	8	9	10	12	11	

6.Below are given the figures of production (in thousand quintal) of a sugar factory.

Year	1985	1987	1988	1989	1990	1991	1994
Production('000 units)	77	88	94	85	91	98	90

Fit a straight line by the 'least squares' method and tabulate the trend values.

7.Explain the utility of time series.

8.The following table relates to the tourist arrivals during 1990 to 1996 in india:

Year	1990	1991	1992	1993	1994	1995	1996
Tourist arrivals (in millions)	18	20	23	25	24	28	30

Fit a straight line trend by the method of least squares and estimate the number of tourists that would arrive in the year 2000.

9.Fit a straight line trend for the following series. Estimate the value for 1997.

Year	1990	1991	1992	1993	1994	1995	1996
Production of Steel (m.tonnes)	60	72	75	65	80	85	95

10.The following data relate to the number of passenger cars in (million) sold from 1992-1999:

Year	1992	1993	1994	1995	1996	1997	1998	1999
Number	6.7	5.3	4.3	6.1	5.6	7.9	5.8	6.1

Fit a straight line trend to the data through 1997 only.

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Post)
Coimbatore –641 021

DEPARTMENT OF MATHEMATICS

SUBJECT: STATISTICS FOR BUSINESS DECISIONS SEMESTER: I L T P C
SUBJECT CODE: 17BAU102 CLASS:I UG (BBA) 5 0 0 5

UNIT IV

Index Numbers: Meaning and significance, problems in construction of index numbers, methods of constructing index numbers-weighted and unweighted, Test of adequacy of index numbers, chain index numbers, base shifting, splicing and deflating index number.

TEXT BOOKS

1. Gupta, S.P. Statistical Methods (34th ed.). New Delhi: Sultan Chand & Sons.

REFERENCES

1. Pillai, R.S.N., & Bagavathi , V. (2002). Statistics . New Delhi: S. Chand & Company Ltd.

2. Srivastava, T N., & Shailaja Rego. (2012). Statistics for Management. New Delhi: Mc Graw Hill Education .

3. Dr. Arora, P.N. (1997). A foundation course statistics. New Delhi: S.chand & Company Ltd.

4.Robert E.stine (2013) Statistics for Business: Decision Making and Analysis : Publisher: Pearson Education; 2 edition (2013)

INDEX NUMBERS

Introduction:

An index number is a statistical device for comparing the general level of magnitude of a group of related variables in two or more situation. If we want to compare the price level of 2000 with what it was in 1990, we shall have to consider a group of variables such as price of wheat, rice, vegetables, cloth, house rent etc., If the changes are in the same ratio and the same direction, we face no difficulty to find out the general price level. But practically, if we think changes in different variables are different and that too, upward or downward, then the price is quoted in different units i.e milk for litre, rice or wheat for kilogram, rent for square feet, etc

We want one figure to indicate the changes of different commodities as a whole. This is called an Index number. Index Number is a number which indicate the changes in magnitudes. M.Spiegel say, " An index number is a statistical measure designed to show changes in variable or a group of related variables with respect to time, geographic location or other characteristic". In general, index numbers are used to measure changes over time in magnitude which are not capable of direct measurement.

On the basis of study and analysis of the definition given above, the following characteristics of index numbers are apparent.

1. Index numbers are specified averages.
2. Index numbers are expressed in percentage.
3. Index numbers measure changes not capable of direct measurement.
4. Index numbers are for comparison.

Uses of Index numbers

Index numbers are indispensable tools of economic and business analysis. They are particular useful in measuring relative changes. Their uses can be appreciated by the following points.

1. They measure the relative change.
2. They are of better comparison.

3. They are good guides.
4. They are economic barometers.
5. They are the pulse of the economy.
6. They compare the wage adjuster.
7. They compare the standard of living.
8. They are a special type of averages.
9. They provide guidelines to policy.
10. To measure the purchasing power of money.

Notation: For any index number, two time periods are needed for comparison. These are called the Base period and the Current period. The period of the year which is used as a basis for comparison is called the base year and the other is the current year. The various notations used are as given below:

P_1 = Price of current year P_0 = Price of base year

q_1 = Quantity of current year q_0 = Quantity of base year

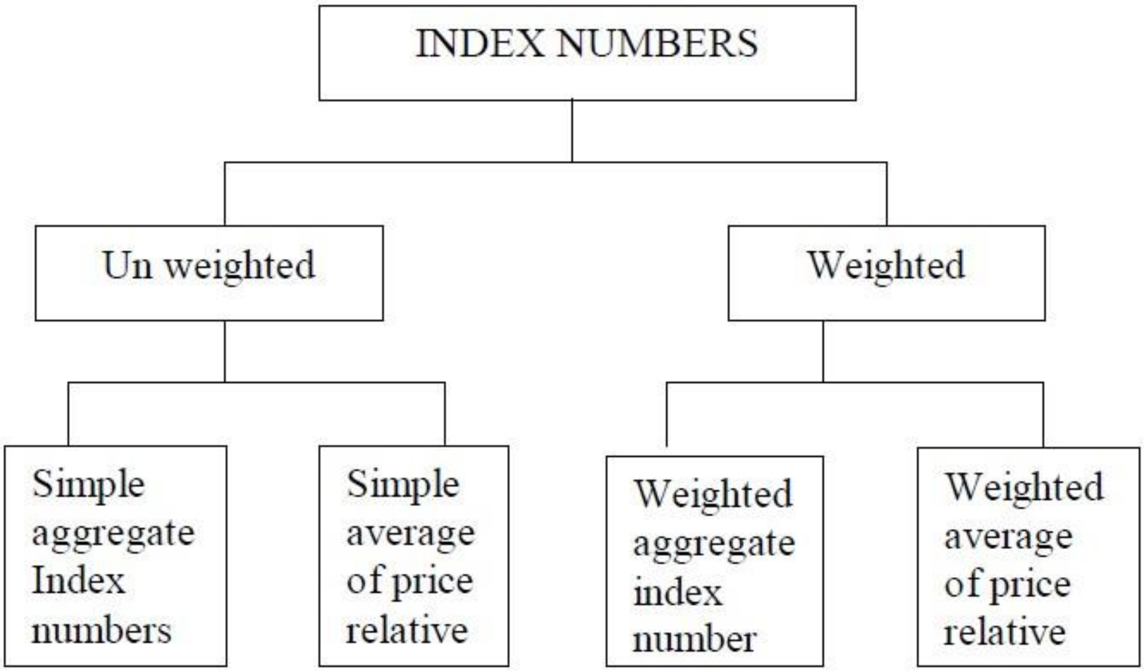
Problems in the construction of index numbers

No index number is an all purpose index number. Hence, there are many problems involved in the construction of index numbers, which are to be tackled by an economist or statistician. They are

1. Purpose of the index numbers
2. Selection of base period
3. Selection of items
4. Selection of source of data
5. Collection of data
6. Selection of average
7. System of weighting

Method of construction of index numbers:

Index numbers may be constructed by various methods as shown below:



Simple Aggregate Index Number

This is the simplest method of construction of index numbers. The price of the different commodities of the current year are added and the sum is divided by the sum of the prices of those commodities by 100. Symbolically,

$$\text{Simple aggregate price index} = P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

Where , Σp_1 = total prices for the current year

Σp_0 = Total prices for the base year

Example 1:

Calculate index numbers from the following data by simple aggregate method taking prices of 2000 as base.

Commodity	Price per unit (in Rupees)	
	2000	2004
A	80	95
B	50	60
C	90	100
D	30	45

Solution:

Commodity	Price per unit (in Rupees)	
	2000 (P_0)	2004 (P_1)
A	80	95
B	50	60
C	90	100
D	30	45
Total	250	300

$$\text{Simple aggregate Price index} = P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

$$= \frac{300}{250} \times 100 = 120$$

Simple Average Price Relative index:

In this method, first calculate the price relative for the various commodities and then average of these relative is obtained by using arithmetic mean and geometric mean. When arithmetic mean is used for average of price relative, the formula for computing the index is

Simple average of price relative by arithmetic mean

$$P_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{n}$$

P_1 = Prices of current year
P_0 = Prices of base year
n = Number of items or commodities
when geometric mean is used for average of price relative, the formula for obtaining the index is
Simple average of price relative by geometric Mean

$$P_{01} = \text{Antilog}\left(\frac{\Sigma \log(\frac{p_1}{p_0} \times 100)}{n}\right)$$

Example 2:
From the following data, construct an index for 1998 taking 1997 as base by the average of price relative using (a) arithmetic mean and (b) Geometric mean

Commodity	Price in 1997	Price in 1998
A	50	70
B	40	60
C	80	100
D	20	30

Solution:
(a) Price relative index number using arithmetic mean

Commodity	Price in 1997 (P_0)	Price in 1998 (P_1)	$\frac{p_1}{p_0} \times 100$
A	50	70	140
B	40	60	150
C	80	100	125
D	20	30	150
		Total	565

$$\text{Simple average of price relative index} = (P_{01}) = \frac{\Sigma\left(\frac{p_1}{p_0}\times 100\right)}{4}$$

$$= \frac{565}{4} = 141.25$$

(b) Price relative index number using Geometric Mean

Commodity	Price in 1997 (P_0)	Price in 1998 (P_1)	$\frac{p_1}{p_0}\times 100$	$\log(\frac{p_1}{p_0}\times 100)$
A	50	70	140	2.1461
B	40	60	150	2.1761
C	80	100	125	2.0969
D	20	30	150	2.1761
			Total	8.5952

Simple average of price Relative index

$$(P_{01}) = \text{Antilog}\ \frac{\Sigma\log\left[\frac{p_1}{p_o}x100\right]}{n}$$

$$= \text{Antilog}\ \frac{8.5952}{4}$$

$$= \text{Antilog}\ [\,2.1488] = 140.9$$

Weighted aggregate index numbers

In order to attribute appropriate importance to each of the items used in an aggregate index number some reasonable weights must be used. There are various methods of assigning weights and consequently a large number of formulae for constructing index numbers have been devised of which some of the most important ones are

1. **Laspeyre's method**
2. **Paasche's method**
3. **Fisher's ideal Method**
4. **Bowley's Method**
5. **Marshall- Edgeworth method**
6. **Kelly's Method**

1. Laspeyre's method:

The Laspeyres price index is a weighted aggregate price index, where the weights are determined by quantities in the based period and is given by

$$\text{Laspeyre's price index} = P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

2. Paasche's method

The Paasche's price index is a weighted aggregate price index in which the weight are determined by the quantities in the current year. The formulae for constructing the index is

$$\text{Paasche's price index number} = P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

Where

P_0 = Price for the base year P_1 = Price for the current year

q_0 = Quantity for the base year q_1 = Quantity for the current year

3. Fisher's ideal Method

Fisher's Price index number is the geometric mean of the Laspeyres and Paasche indices Symbolically

$$\text{Fisher's ideal index number} = P_{01}^{F} = \sqrt{L \times P}$$

$$= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

It is known as ideal index number because

(a) It is based on the geometric mean
(b) It is based on the current year as well as the base year
(c) It conform certain tests of consistency
(d) It is free from bias.

4. Bowley's Method:

Bowley's price index number is the arithmetic mean of Laspeyre's and Paasche's method. Symbolically

$$\text{Bowley's price index number} = P_{01}^{B} = \frac{L + P}{2}$$

$$= \frac{1}{2}\left[\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right] \times 100$$

5. Marshall- Edgeworth method

This method also both the current year as well as base year prices and quantities are considered. The formula for constructing the index is

$$\text{Marshall Edgeworth price index} = P_{01}^{ME} = \frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} \times 100$$

$$= \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

6. Kelly' s Method

Kelly has suggested the following formula for constructing the index number

$$\text{Kelly' s Price index number} = P_{01}^{k} = \frac{\Sigma p_1 q}{\Sigma p_0 q} \times 100$$

$$\text{Where } = q = \frac{q_0 + q_1}{2}$$

Here the average of the quantities of two years is used as weights

Example 3:

Construct price index number from the following data by applying

1. Laspeyere' s Method
2. Paasche' s Method
3. Fisher' s ideal Method

Commodity	2000		2001	
	Price	Qty	Price	Qty
A	2	8	4	5
B	5	12	6	10
C	4	15	5	12
D	2	18	4	20

Solution:

Commodity	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	$p_1 q_1$
A	2	8	4	5	16	10	32	20
B	5	12	6	10	60	50	72	60
C	4	15	5	12	60	48	75	60
D	2	18	4	20	36	40	72	80
					172	148	251	220

$$\text{Laspeyre's price index} = P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{251}{172} \times 100 \quad = 145.93$$

$$\text{Paasche price index number} = P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{220}{148} \times 100$$

$$= 148.7$$

$$\text{Fisher's ideal index number} = \sqrt{L \times P}$$

$$= \sqrt{(145.9) \times (148.7)}$$

$$= \sqrt{21695.33}$$

$$= 147.3$$

Or

$$\text{Fisher's ideal index number} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{251}{172} \times \frac{220}{148}} \times 100$$

$$= \sqrt{(1.459) \times (1.487)} \times 100$$

$$= \sqrt{2.170} \times 100$$

$$= 1.473 \times 100 = 147.3$$

Interpretation:

The results can be interpreted as follows:

If 100 rupees were used in the base year to buy the given commodities, we have to use Rs 145.90 in the current year to buy the same amount of the commodities as per the Laspeyre's formula. Other values give similar meaning .

Example 4:

Calculate the index number from the following data by applying

(a) Bowley's price index

(b) Marshall- Edgeworth price index

Commodity	Base year		Current year	
	Quantity	Price	Quantity	Price
A	10	3	8	4
B	20	15	15	20
C	2	25	3	30

Solution:

Commodity	q_0	P_0	q_1	P_1	p_0q_0	p_0q_1	p_1q_0	$p_1 q_1$
A	10	3	8	4	30	24	40	32
B	20	15	15	20	300	225	400	300
C	2	25	3	30	50	75	60	90
					380	324	500	422

(a) Bowley' s price index number $= \frac{1}{2}\left[\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right] \times 100$

$$= \frac{1}{2}\left[\frac{500}{380} + \frac{422}{324}\right] \times 100$$

$$= \frac{1}{2}\left[1.316 + 1.302\right] \times 100$$

$$= \frac{1}{2}\left[2.168\right] \times 100$$

$$= 1.309 \times 100$$

$$= 130.9$$

(b) Marshall Edgeworths price index Number

$$= P_{01}^{ME} = \frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} \times 100$$

$$= \left[\frac{500}{380} + \frac{422}{324}\right] \times 100$$

$$= \left[\frac{922}{704}\right] \times 100$$

$= 131.0$

Example 5:

Calculate a suitable price index from the following data

Commodity	Quantity	Price	
		1996	1997
A	20	2	4
B	15	5	6
C	8	3	2

Solution:

Here the quantities are given in common we can use Kelly's index price number and is given by

$$\text{Kelly's Price index number} = P_{01}^{k} = \frac{\Sigma p_1 q}{\Sigma p_0 q} \times 100$$

$$= \frac{186}{139} \times 100 = 133.81$$

Commodity	q	P_0	P_1	p_0q	P_1q
A	20	2	4	40	80
B	15	5	6	75	90
C	8	3	2	24	16
			Total	139	186

$$\text{Kelly's Price index number} = P_{01}^{k} = \frac{\Sigma p_1 q}{\Sigma p_0 q} \times 100$$

IV. Weighted Average of Price Relative index.

When the specific weights are given for each commodity, the weighted index number is calculated by the formula.

$$\text{Weighted Average of Price Relative index} = \frac{\Sigma pw}{\Sigma w}$$

Where w = the weight of the commodity

P = the price relative index

$$= \frac{p_1}{p_0} \times 100$$

When the base year value P_0q_0 is taken as the weight i.e. $W=P_0q_0$ then the formula is

$$\text{Weighted Average of Price Relative index} = \frac{\Sigma\left(\frac{p_1}{p_0}\times 100\right)\times p_0q_0}{\Sigma p_0q_0}$$

$$= \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$$

This is nothing but Laspeyre' s formula.

When the weights are taken as $w = p_0q_1$, the formula is

$$\text{Weighted Average of Price Relative index} = \frac{\Sigma\left(\frac{p_1}{p_0}\times 100\right)\times p_0q_1}{\Sigma p_0q_1}$$

$$= \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$$

This is nothing but Paasche' s Formula.

Example 6:

Compute the weighted index number for the following data.

Commodity	Price		Weight
	Current year	Base year	
A	5	4	60
B	3	2	50
C	2	1	30

Solution:

Commodity	P_1	P_0	W	$P= \frac{p_1}{p_0} \times 100$	PW
A	5	4	60	125	7500
B	3	2	50	150	7500

C	2	1	30	200	6000
			140		21000

Weighted Average of Price Relative index = $\frac{\Sigma pw}{\Sigma w}$

$$= \frac{21000}{140}$$

$$= 150$$

Test of adequacy:

Several formulae have been studied for the construction of index number. The question arises as to which formula is appropriate to a given problems. A number of tests been developed and the important among these are

1. Unit test
2. Time Reversal test
3. Factor Reversal test

1. Unit test:

The unit test requires that the formula for constructing an index should be independent of the units in which prices and quantities are quoted. Except for the simple aggregate index (unweighted) , all other formulae discussed in this chapter satisfy this test.

2. Time Reversal test:

Time Reversal test is a test to determine whether a given method will work both ways in time, forward and backward. In the words of Fisher, "the formula for calculating the index number should be such that it gives the same ratio between one point of comparison and the other, no matter which of the two is taken as base". Symbolically, the following relation should be satisfied.

$$P_{01} \times P_{10} = 1$$

Where P_{01} is the index for time ' 1' as time ' 0' as base and P_{10} is the index for time ' 0' as time ' 1' as base. If the product is not unity, there is said to be a time bias is the method. Fisher' s ideal index satisfies the time reversal test.

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$\text{Then } P_{01} \times P_{10} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$= \sqrt{1} = 1$$

Therefore Fisher ideal index satisfies the time reversal test.

3. Factor Reversal test:

Another test suggested by Fisher is known s factor reversal test. It holds that the product of a price index and the quantity index should be equal to the corresponding value index. In the words of Fisher, "Just as each formula should permit the interchange of the two times without giving inconsistent results, so it ought to permit interchanging the prices and quantities without giving inconsistent result, ie, the two results multiplied together should give the true value ratio.

In other word, if P_{01} represent the changes in price in the current year and Q_{01} represent the changes in quantity in the current year, then

$$P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Thus based on this test, if the product is not equal to the value ratio, there is an error in one or both of the index number. The Factor reversal test is satisfied by the Fisher' s ideal index.

ie. $$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

Then $$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \sqrt{\left(\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}\right)^2}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Since $P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$, the factor reversal test is satisfied by the Fisher' s ideal index.

Example 8:
Construct Fisher' s ideal index for the Following data. Test whether it satisfies time reversal test and factor reversal test.

	Base year		Current year	
Commodity	Quantity	Price	Quantity	Price
A	12	10	15	12
B	15	7	20	5
C	5	5	8	9

Solution:

Commodity	q_0	p_0	q_1	p_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	12	10	15	12	120	150	144	180
B	15	7	20	5	105	140	75	100
C	5	5	8	9	25	40	45	72
					250	330	264	352

Fisher ideal index number $P_{01}^F = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$

$$= \sqrt{\frac{264}{250} \times \frac{352}{330}} \times 100$$

$$= \sqrt{(1.056) \times (1.067)} \times 100$$

$$= \sqrt{1.127} \times 100$$

$$= 1.062 \times 100 = 106.2$$

Time Reversal test:

Time Reversal test is satisfied when $P_{01} \times P_{10} = 1$

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$$

$$= \sqrt{\frac{264}{250} \times \frac{352}{330}}$$

$$P_{10} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}}$$

$$= \sqrt{\frac{330}{352} \times \frac{250}{264}}$$

Now $P_{01} \times P_{10} = \sqrt{\frac{264}{250} \times \frac{352}{330} \times \frac{330}{352} \times \frac{250}{264}}$

$$= \sqrt{1}$$

$$= 1$$

Hence Fisher ideal index satisfy the time reversal test.

Factor Reversal test:

Factor Reversal test is satisfied when $P_{01} \times Q_{01} = \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

Now $P_{01} = \sqrt{\dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$

$= \sqrt{\dfrac{264}{250} \times \dfrac{352}{330}}$

$Q_{01} = \sqrt{\dfrac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \dfrac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$

$= \sqrt{\dfrac{330}{250} \times \dfrac{352}{264}}$

Then $P_{01} \times Q_{01} = \sqrt{\dfrac{264}{250} \times \dfrac{352}{330} \times \dfrac{330}{250} \times \dfrac{352}{264}}$

$= \sqrt{\left(\dfrac{352}{250}\right)^2}$

$= \dfrac{352}{250}$

$= \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

Hence Fisher ideal index number satisfy the factor reversal test.

BASE SHIFTING

The need for shifting the base may arise either

(i) when the base period of a given index number series is to be made more recent,

or

(ii) when two index number series with different base periods are to be compared,

or

(iii) when there is need for splicing two overlapping index number series.

Whatever be the reason, the technique of shifting the base is simple:

New Base Index Number $\frac{\text{Old index number of new base year}}{\text{Old index number of current year}}$ x 100

Example

Reconstruct the following indices using 1997 as base:

Year : 1991 1992 1993 1994 1995 1996 1997 1998
Index : 100 110 130 150 175 180 200 220

Solution:
Shifting the Base Period

Year	Index Number (1991 = 100)	Index Number (1997 = 100)
1991	100	(100/200) *x*100 = 50.00
1992	110	(110/200) *x*100 = 55.00
1993	130	(130/200) *x*100 = 65.00
1994	150	(150/200) *x*100 = 75.00
1995	175	(175/200) *x*100 = 87.50
1996	180	(180/200) *x*100 = 90.00
1997	200	(200/200) *x*100 = 100.00
1998	220	(220/200) *x*100 = 110.00

SPLICING TWO OVERLAPPING INDEX NUMBER SERIES

Splicing two index number series means reducing two overlapping index series with different base periods into a single series either at the base period of the old series (one with an old base year), or at the base period of the new series (one with a recent base year). This actually amounts to changing the weights of one series into the weights of the other series.

1. Splicing the New Series to Make it Continuous with the Old Series

Here we reduce the new series into the old series after the base year of the former. As shown in Table 6.8.2*(i)*, splicing here takes place at the base year (1980) of the new series. To do this, a ratio of the index for 1980 in the old series (200) to the index of 1980 in the new series (100) is computed and the index for each of the following years in the new series is multiplied by this ratio.

Year	Price Index (1976 = 100) (Old Series)	Price Index (1980 = 100) (New Series)	Spliced Index Number [New Series *x* (200/100)]
1976	100	--	100
1977	120	--	120
1978	146	--	146
1979	172	--	172
1980	200	100	200
1981	--	110	220
1982	--	116	232
1983	--	125	250
1984	--	140	280

Splicing the Old Series to Make it Continuous with the New Series

This means reducing the old series into the new series before the base period of the letter. As shown in Table 6.8.2*(ii),* splicing here takes place at the base period of the new series. To do this, a ratio of the index of 1980 of the new series (100) to the index of 1980 of the old series (200) is computed and the index for each of the preceding years of the old series are then multiplied by this ratio.

Year	Price Index (1976 = 100) (Old Series)	Price Index (1980 = 100) (New Series)	Spliced Index Number [Old Series *x* (100/200)]
1976	100	--	50
1977	120	--	60
1978	146	--	73.50
1979	172	--	86
1980	200	100	100
1981	--	110	110
1982	--	116	116
1983	--	125	125
1984	--	140	140

Deflating

It is a technique used to make allowances for the effect of changing price values. It is used to measure the purchasing power of money.

$$\text{Deflated value} = \frac{\text{Current Value}}{\text{Price index of the current year}} \times 100$$

It can also be found using the relation

$$\text{Deflated value} = \text{Current value} \times \frac{\text{Base price } (p_0)}{\text{Current price } (p_1)}$$

Example

The table below shows the income of a company and its price index taking the year 1991 as the base.

Years	1991	1992	1993	1994	1995	1996
Income (in crores)	80	108	125	147	216	230
Price Index	100	120	125	140	180	200

Calculate the deflated value for every year taking 1991 as the base.

Solution

Years	Income (in crores)	Price Index	Deflated Value
1991	80	100	$\frac{\text{Current Value}}{\text{Price index of the current year}} \times 100 = \frac{80}{100} \times 100 = 80$
1992	108	120	$\frac{108}{120} \times 100 = 90$
1993	125	125	$\frac{125}{125} \times 100 = 100$
1994	147	140	$\frac{147}{140} \times 100 = 105$
1995	216	180	$\frac{216}{180} \times 100 = 120$
1996	230	200	$\frac{230}{200} \times 100 = 115$

Possible questions

TWO MARKS:

1. Define deflating.
2. Define base shifting.
3. Define Splicing.
4. Write the characteristics of index numbers?

EIGHT MARKS:

1.Calculate the weighted price index from the following data:

Materials Required	Unit	Quantity Required	Price during 1973 Rs.	1977 Rs.
Cement	50 kgs	250 kgs	50	80
Timber	m^3	70 m^3	300	400
Steel Sheet	Quintals	5 quintals	340	420
Bricks	per’000	20,000	120	240

2.Obtain Laspeyre’s price index number and Paasche’s quantity index number from following data:

	Price (Rs. per unit)		Quantity	
Item	Base year	Current year	Base year	Current year
1	2	5	20	15
2	4	8	4	5
3	1	2	10	12
4	5	10	5	6

3.The following table gives the prices and quantities of four commodities A,B,Cand D for the years 1985 and 1988. Calculate the price index for 1988 and 1985 as base by using Marshall-Edgeworth’s method. Compare this index with Laspeyre’s index number.

Commoditiy	1985		1988	
	Price	Qty.	Price	Qty.
A	40	10	50	7
B	20	5	30	8

C	30	6	40	10
D	10	9	20	10

4.The following are the index numbers of prices (1990=100):

Year :	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Index:	100	110	120	200	400	410	400	380	370	340

Shift the base from 1990 to 1996 and recast the index numbers.

5.What are the problems in the construction of index numbers

6.It is stated that Marshall-Edgeworth index number is a good approximation to the ideal index number. Verify using the following data:

Commoditiy	1994		1995	
	Price	Qty.	Price	Qty.
A	2	74	3	82
B	5	125	4	140
C	7	40	6	33

7.Explain the test of adequacy of index numbers.

8.The index A given was started in 1982 and continued upto 1992 in which year another index B was started. Splice the index B to index A o that a continuous series of index

Year	Index A	Index B	Year	Index A	Index B
1982	100		1991	138	
1983	110		1992	150	100
1984	112		1993		120
---			1994		140
---			1995		130
---			1996		150

9.From the following data construct the index number for the year 1992 taking 1991 as Base

by using (i) Arithmetic Mean (ii) Geometric Mean.

Item	Price Rs. (1991)	Price Rs. (1992)
A	6	10
B	2	2
C	4	6
D	10	12
E	8	12

10.Construct Fisher's Ideal Index Number for the following data and show how it satisfies the Time and Factor Reversal Tests.

Commodities	1998		1999	
	Qty.	Price	Qty.	Price
M	20	12	30	14
N	13	14	15	20
O	12	10	20	15
P	8	6	10	4
Q	5	8	5	6

KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Post)
Coimbatore –641 021

DEPARTMENT OF MATHEMATICS

SUBJECT: STATISTICS FOR BUSINESS DECISIONS	**SEMESTER: I**	**L T P C**
SUBJECT CODE: 17BAU102	**CLASS:I UG (BBA)**	**5 0 0 5**

UNIT V

Probability: Meaning and need. Theorems of addition and multiplication. Conditional probability. Bayes' theorem, Random Variable- discrete and continuous. Probability Distribution: Meaning, characteristics (Expectation and variance) of Binomial, Poisson, and Normal distribution. Central limit theorem.

TEXT BOOKS

1. Gupta, S.P. Statistical Methods (34th ed.). New Delhi: Sultan Chand & Sons.

2. Richard Levin & David Rubin . Statistics for management. New Delhi: Prentice Hall.

REFERENCES

1. Pillai, R.S.N., & Bagavathi , V. (2002). Statistics . New Delhi: S. Chand & Company Ltd

2. Dr. Arora, P.N. (1997). A foundation course statistics. New Delhi: S.chand & Company Ltd.

PROBABILITY

Random experiment:

Random experiment is one whose results depend on chance, that is the result cannot be predicted. Tossing of coins, throwing of dice are some examples of random experiments.

Trial:

Performing a random experiment is called a trial.

Outcomes:

The results of a random experiment are called its outcomes. When two coins are tossed the possible outcomes are HH, HT, TH, TT.

Probability theory is being applied in the solution of social, economic, business problems. Today the concept of probability has assumed greater importance and the mathematical theory of probability has become the basis for statistical applications in both social and decision-making research. Probability theory, in fact, is the foundation of statistical inferences.

Interpretation of statistical statements in terms of set theory:

$S \Rightarrow$ Sample space

$\overline{A} \Rightarrow$ A does not occur

$A \cup \overline{A} = S$

$A \cap B = \phi \Rightarrow$ A and B are mutually exclusive.

$A \cup B \Rightarrow$ Event A occurs or B occurs or both A and B occur. (at least one of the events A or B occurs)

$A \cap B \Rightarrow$ Both the events A and B occur.

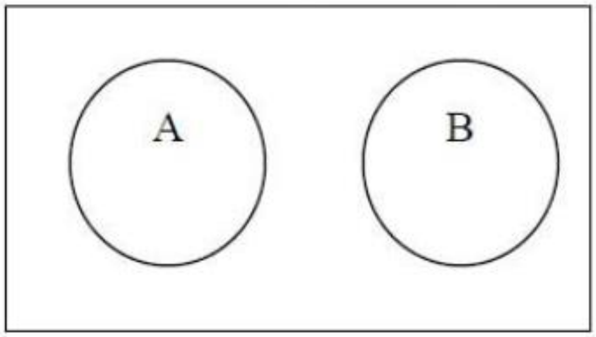
$\overline{A} \cap \overline{B} \Rightarrow$ Neither A nor B occurs

$A \cap \overline{B} \Rightarrow$ Event A occurs and B does not occur

$\overline{A} \cap B \Rightarrow$ Event A does not occur and B occur.

Addition theorem on probabilities:

If two events A and B are mutually exclusive, the probability of the occurrence of either A or B is the sum of individual probabilities of A and B. ie $P(A\cup B) = P(A) + P(B)$
This is clearly stated in axioms of probability.



Addition theorem on probabilities for not-mutually exclusive events:

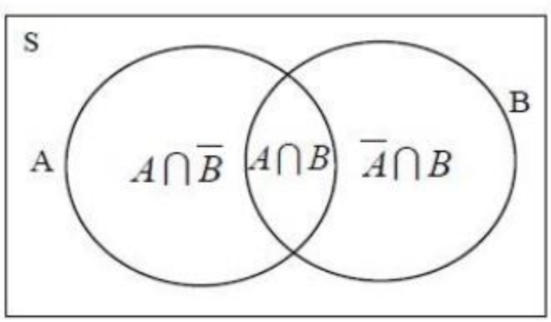
If two events A and B are not-mutually exclusive, the probability of the event that either A or B or both occur is given as

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

Proof:

Let us take a random experiment with a sample space S of N sample points.
Then by the definition of probability ,

$$P(A\cup B) = \frac{n(A\cup B)}{n(S)} = \frac{n(A\cup B)}{N}$$



From the diagram, using the axiom for the mutually exclusive events, we write

$$P(AUB) = \frac{n(A) + n(\overline{A} \cap B)}{N}$$

Adding and subtracting n($A \cap B$) in the numerator,

$$= \frac{n(A) + n(\overline{A} \cap B) + n(A \cap B) - n(A \cap B)}{N}$$

$$= \frac{n(A) + n(B) - n(A \cap B)}{N}$$

$$= \frac{n(A)}{N} + \frac{n(B)}{N} - \frac{n(A \cap B)}{N}$$

$$P(AUB) = P(A) + P(B) - P(A \cap B)$$

Conditional probability:

Let A be any event with p(A) >0. The probability that an event B occurs subject to the condition that A has already occurred is known as the conditional probability of occurrence of the event B on the assumption that the event A has already occurred and is denoted by the symbol P(B/A) or P(B|A) and is read as the probability of B given A.

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

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

Similarly the conditional probability of A given B is given as

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

BAYES' Theorem:

Bayes' Theorem or Rule (Statement only):

Let $A_1, A_2, A_3, \ldots A_i, \ldots A_n$ be a set of n mutually exclusive and collectively exhaustive events and $P(A_1), P(A_2) \ldots P(A_n)$ are their corresponding probabilities. If B is another event such that P(B) is not zero and the priori probabilities $P(B|A_i)$ $i = 1, 2, \ldots, n$ are also known. Then

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{\sum_{i=1}^{k} P(B \mid A_i)\, P(A_i)}$$

EXAMPLE:

A box contains 5 red and 4 white marbles. Two marbles are drawn successively from the box without replacement and it is noted that the second one is white. What is the probability that the first is also white?

Solution:

If w_1, w_2 are the events ' white on the first draw' , ' white on the second draw' respectively.

Now we are looking for $P(w_1/w_2)$

$$P(w_1/w_2) = \frac{P(w_1 \cap w_2)}{P(w_2)} = \frac{P(w_1).P(w_2)}{P(w_2)}$$

$$= \frac{(4/9)(3/8)}{(3/8)}$$

$$= \frac{4}{9}$$

EXAMPLE:

A bag contains 6 red and 8 black balls. Another bag contains 7 red and 10 black balls. A bag is selected and a ball is drawn. Find the probability that it is a red ball.

Solution:

There are two bags

$\therefore$ probability of selecting a bag = $\frac{1}{2}$

Let A denote the first bag and B denote the second bag.

Then P(A) = P(B)= $\frac{1}{2}$

Bag ‘ A’ contains 6 red and 8 black balls.

$\therefore$ Probability of drawing a red ball is $\frac{6}{14}$

Probability of selecting bag A and drawing a red ball from that bag is P(A). P(R/A) = $\frac{1}{2} \times \frac{6}{14} = \frac{3}{14}$

Similarly probability of selecting bag B and drawing a red ball from that bag is P(B). P(R/B) = $\frac{1}{2} \times \frac{7}{17} = \frac{7}{34}$

All these are mutually exclusive events

∴ Probability of drawing a red ball either from the bag A or B is

P(R) = P(A) P(R/A) + P(B) P(R/B)

$$= \frac{3}{14} + \frac{7}{34}$$

$$= \frac{17 \times 3 + 7 \times 7}{238}$$

$$= \frac{51 + 49}{238}$$

$$= \frac{100}{238} = \frac{50}{119}$$

BINOMIAL DISTRIBUTION

A random variable X is said to follow binomial distribution, if its probability mass function is given by

$$P(X = x) = P(x) = \begin{cases} nC_x \; p^x q^{n-x} & ; \quad x = 0, 1, 2, \ldots, n \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Characteristics of Binomial Distribution:

1. Binomial distribution is a discrete distribution in which the random variable X (the number of success) assumes the values 0,1, 2, ...n, where n is finite.
2. Mean = np, variance = npq and standard deviation $\sigma = \sqrt{npq}$,

   Coefficient of skewness $= \dfrac{q-p}{\sqrt{npq}}$,

Example 1:

Comment on the following: " The mean of a binomial distribution is 5 and its variance is 9"

Solution:

The parameters of the binomial distribution are n and p

We have mean $\Rightarrow$ np = 5

Variance $\Rightarrow$ npq = 9

$$\therefore q = \frac{npq}{np} = \frac{9}{5}$$

$$q = \frac{9}{5} > 1$$

Which is not admissible since q cannot exceed unity. Hence the given statement is wrong.

Example 2:

Eight coins are tossed simultaneously. Find the probability of getting atleast six heads.

Solution:

Here number of trials, n = 8, p denotes the probability of getting a head.

$$\therefore p = \frac{1}{2} \text{ and } q = \frac{1}{2}$$

If the random variable X denotes the number of heads, then the probability of a success in n trials is given by

$$P(X = x) = nc_x\, p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

$$= 8C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{8-x} = 8C_x \left(\frac{1}{2}\right)^8$$

$$= \frac{1}{2^8}\, 8C_x$$

Probability of getting atleast six heads is given by

$$P(x \geq 6) = P(x = 6) + P(x = 7) + P(x = 8)$$

$$= \frac{1}{2^8}\, 8C_6 + \frac{1}{2^8}\, 8C_7 + \frac{1}{2^8}\, 8C_8$$

$$= \frac{1}{2^8}\, [\, 8C_6 + 8C_7 + 8C_8]$$

$$= \frac{1}{2^8}\, [\, 28 + 8 + 1] = \frac{37}{256}$$

POISSON DISTRIBUTION:

The probability of x success is given by

$$P(X = x) = \begin{cases} \dfrac{e^{-m} m^x}{x!} & \text{for } x = 0, 1, 2, \ldots \\ 0 & ; \text{ otherwise} \end{cases}$$

Here m is known as parameter of the distribution so that m >0

Note:

1) e is given by $e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots = 2.71828$
2) $P(X=0) = \frac{e^{-m} m^0}{0!}$, $0! = 1$ and $1! = 1$
3) $P(X=1) = \frac{e^{-m} m^1}{1!}$

EXAMPLE:

If 2% of electric bulbs manufactured by a certain company are defective. Find the probability that in a sample of 200 bulbs i) less than 2 bulbs ii) more than 3 bulbs are defective.[$e^{-4} = 0.0183$]

Solution:

The probability of a defective bulb $= p = \frac{2}{100} = 0.02$

Given that n = 200 since p is small and n is large
We use the Poisson distribution
mean, $m = np = 200 \times 0.02 = 4$

Now, Poisson Probability function, $P(X = x) = \frac{e^{-m} m^x}{x!}$

i) Probability of less than 2 bulbs are defective

$$= P(X<2)$$
$$= P(x = 0) + P(x = 1)$$
$$= \frac{e^{-4} 4^0}{0!} + \frac{e^{-4} 4^1}{1!}$$
$$= e^{-4} + e^{-4} (4)$$
$$= e^{-4} (1 + 4) = 0.0183 \times 5$$
$$= 0.0915$$

ii) Probability of getting more than 3 defective bulbs

$$P(x > 3) = 1 - P(x \leq 3)$$
$$= 1 - \{P(x = 0) + P(x = 1) + P(x=2) + P(x=3)\}$$
$$= 1 - e^{-4} \{1 + 4 + \frac{4^2}{2!} + \frac{4^3}{3!}\}$$
$$= 1 - \{0.0183 \times (1 + 4 + 8 + 10.67)\}$$
$$= 0.567$$

NORMAL DISTRIBUTION:

Definition:

A continuous random variable X is said to follow normal distribution with mean μ and standard deviation σ, if its probability density function

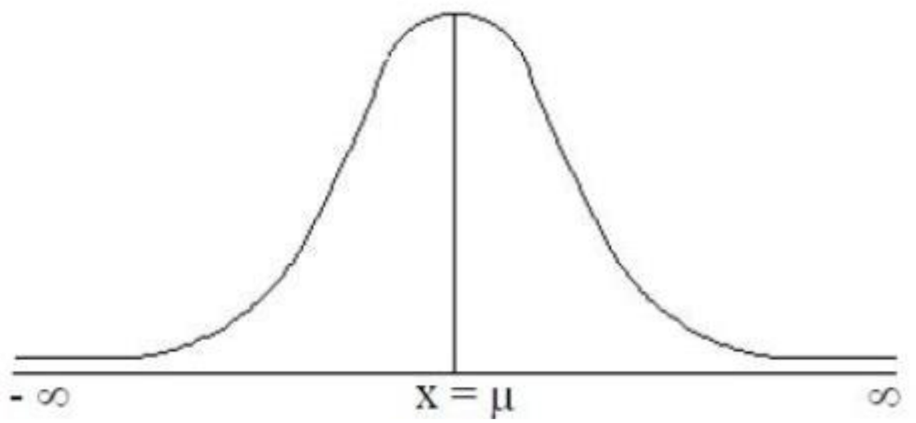
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; -\infty < x < \infty \ , \ -\infty < \mu < \infty, \ \sigma > 0.$$

Note:

The mean μ and standard deviation σ are called the parameters of Normal distribution. The normal distribution is expressed by $X \sim N(\mu, \sigma^2)$

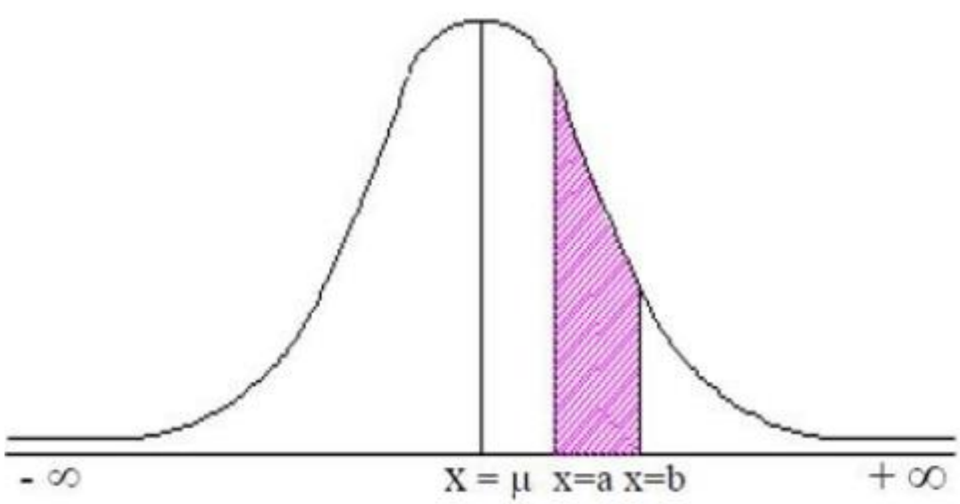
Normal probability curve:

The curve representing the normal distribution is called the normal probability curve. The curve is symmetrical about the mean (μ), bell-shaped and the two tails on the right and left sides of the mean extends to the infinity. The shape of the curve is shown in the following figure.

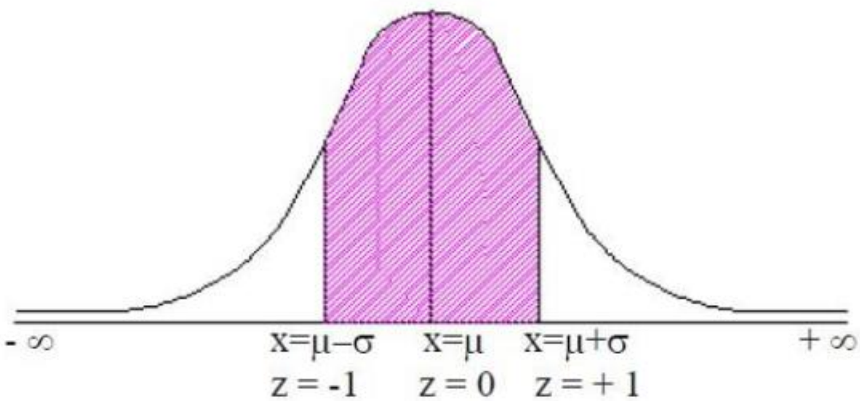


Area properties of Normal curve:

The total area under the normal probability curve is 1. The curve is also called standard probability curve. The area under the curve between the ordinates at $x = a$ and $x = b$ where $a < b$, represents the probabilities that x lies between $x = a$ and $x = b$ i.e., $P(a \le x \le b)$

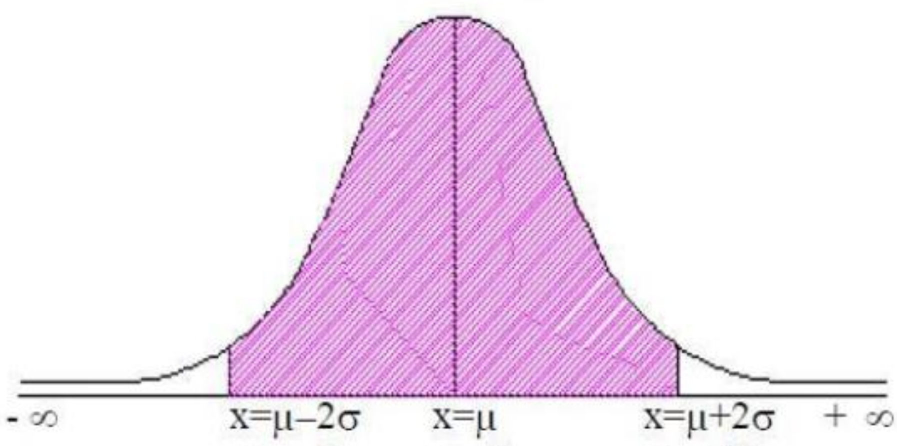


For Example: The probability that the normal random variable x to lie in the interval $(\mu-\sigma, \mu+\sigma)$ is given by

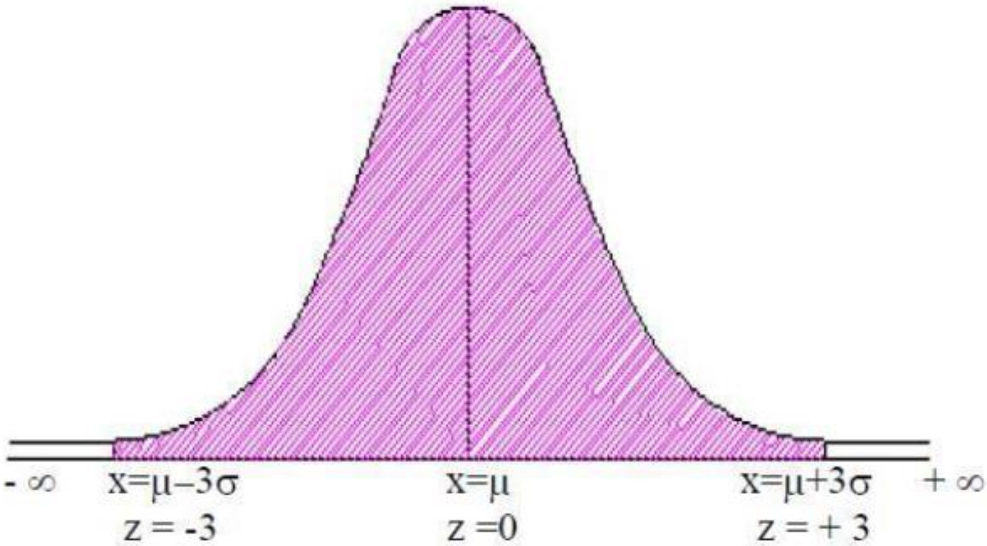


$$P(\mu - \sigma < x < \mu+\sigma) = P(-1 \le z \le 1)$$
$$= 2P(0 < z < 1)$$
$$= 2\,(0.3413) \quad \text{(from the area table)}$$
$$= 0.6826$$

$$P(\mu - 2\sigma < x < \mu+2\sigma) = P(-2 < z < 2)$$
$$= 2P(0 < z < 2)$$
$$= 2\,(0.4772) = 0.9544$$



$$P(\mu - 3\sigma < x < \mu + 3\sigma) = P(-3 < z < 3)$$
$$= 2P(0 < z < 3)$$
$$= 2\,(0.49865) = 0.9973$$



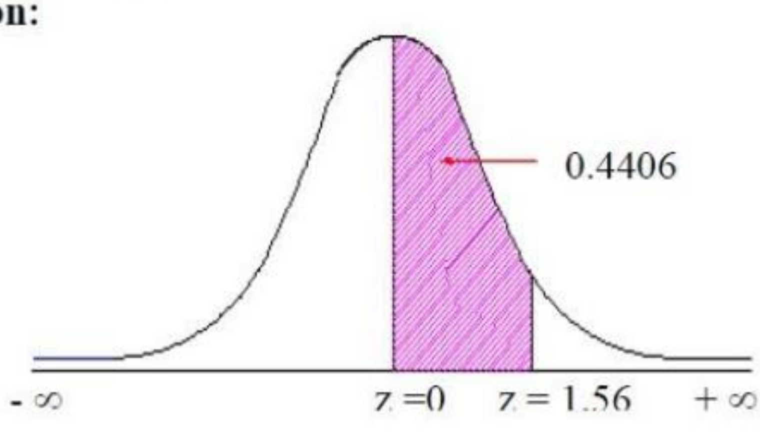
The probability that a normal variate x lies outside the range $\mu \pm 3\sigma$ is given by

$$P(|x - \mu| > 3\sigma) = P(|z| > 3)$$
$$= 1 - P(-3 \le z \le 3)$$
$$= 1 - 0.9773 = 0.0027$$

Thus we expect that the values in a normal probability curve will lie between the range $\mu \pm 3\sigma$, though theoretically it range from $-\infty$ to ∞.

Find the probability that the standard normal variate lies between 0 and 1.56

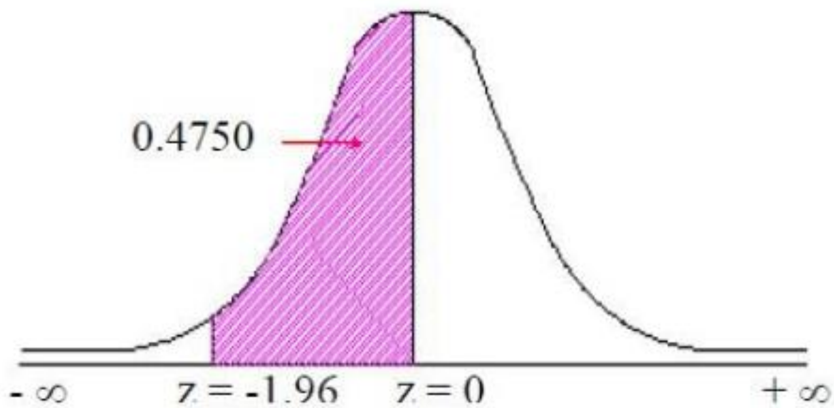
Solution:



$P(0<z<1.56) =$ Area between z = 0 and z = 1.56
$= 0.4406$ (from table)

Find the area of the standard normal variate from –1.96 to 0.

Solution:



Area between z = 0 & z =1.96 is same as the area z = –1.96 to z = 0

$P(-1.96 < z < 0) = P(0 < z < 1.96)$ (by symmetry)

$= 0.4750$ (from the table)

Example

X is normal distribution with mean 2 and standard deviation 3. Find the value of the variable x such that the probability of the interval from mean to that value is 0.4115

Solution:

Given $\mu = 2$, $\sigma = 3$

Suppose z_1 is required standard value,

Thus $P(0 < z < z_1) = 0.4115$

From the table the value corresponding to the area 0.4115 is 1.35 that is $z_1 = 1.35$

$$\text{Here } z_1 = \frac{x-\mu}{\sigma}$$

$$1.35 = \frac{x-2}{3}$$

$$x = 3(1.35) + 2$$

$$= 4.05 + 2 = 6.05$$

CENTRAL LIMIT THEOREM

The central limit theorem explains why the normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

is prevalent. If we add independent random variables and normalize them so that the mean is zero and the standard deviation is 1, then the distribution of the sum converges to the normal distribution.

POSSIBLE QUESTIONS (TWO MARKS)

1. Define deflating.
2. Define base shifting.
3. Define Splicing.
4. Write the characteristics of index numbers?

POSSIBLE QUESTIONS (EIGHT MARKS)

1.A person is known to hit the target in 3 out of 4 shots, whereas another person is known to hit the target in 2 out of 3 shot. Find the probability of the target being hit at all when they both try.

2.Is there any inconsistency in the statement, the mean of binomial distribution is 20 and itstandard deviation 4? If no inconsistency is found what shall be the values of p, q and n.

3.What is the probability of picking a card that was red or black?

4.Three horses A, B and C are in a race. A is twice as likely to win as B and B is as likely to win as C. What are the respective probabilities of winning?

5.Find the probability that he value of an item drawn at random from a normal distribution with mean 20 and standard deviation 10 will be between:

(a) 10 and 15 (b) -5 and 10 (c) 15 and 25

The relevant extract of the area table:

0.5	1.0	1.5	2.0	2.5
0.1915	0.3413	0.4332	0.4772	0.4938

6.A bag contains 5 white and 3 black balls. Two balls are drawn at random one after the other without replacement. Find the probability that both balls drawn are black.

7.In a town 10 accidents took place in a span of 50 days. Assume that the number of accidents per day follows the Poisson distribution; find the probability that there will be three or more accidents in a day.

8.Find the probability that at most 5 defective bolts will be found in a box of 200 bolts, if it is known that 2 % of such bolts are expected to be defective.($e^{-4} = 0.0183$).

9.12 coins are tossed. What are the probabilities in a single toss for getting,

i) 9 or more heads

ii) less than 3 heads

iii) atleast 8 heads

10.A bag contains 10 white and 6 black balls. 4 balls are successively drawn out ant not replaced. What is the probability that they are alternately of different colours?

11.The incidence of occupational disease in an industry is such that the workmen have a 20% chance of suffering from it. What is the probability that out of six workmen, 4 or more will contact the disease?

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Po),
Coimbatore –641 021
DEPARTMENT OF MATHEMATICS
PART-A Multiple Choice Questions (Each Question Carries One Mark)

Subject Name: Statistics for Business Decisions **Subject Code: 17BAU102**

UNIT-I

Question	Option-1	Option-2	Option-3	Option-4	Answer
Which one of the following is a measure of central tendency?	Median	range	variation	correlation	Median
The total of the values of the items divided by their number of items is known as	Median	Arithmetic mean	mode	range	Arithmetic mean
In the short-cut method of arithmetic mean, the deviation is taken as	x – A	A – x	(x – A) / c	(A – x) / c	x – A
The sum of the deviations of the values from their arithmetic mean is	– 1	one	two	zero	zero
The formula for the weighted arithmetic mean is	∑wx / ∑w	∑w / ∑wx	∑x / n	∑x / ∑f	∑wx / ∑w
Find the Mean of the following values. 5, 15, 20, 10, 40	5	18	41	20	18
Which of the followings represents median?	First quartile	Third quartile	Second quartile	Q.D	Second quartile
Which of the measure of central tendency is not affected by extreme values?	Mode	Median	sixth deciles	Mean	Median
Which one of the following is relative measure of dispersion?	Range	Q.D	S.D	coefficient of variation	coefficient of
Quartile deviation is half of the difference between the -------------- quartiles	Q_3 and Q_1	Q_2 and Q_1	Q_4 and Q_1	Q3 and Q2	Q_3 and Q_1
The coefficient of Quartile deviation is given by	(Q3 – Q1)/(Q3+Q1)	(Q3 + Q1)/(Q3-Q1)	(Q3 – Q1)/(Q3-Q1)	(Q3 – Q1)	(Q3 – Q1)/(Q3+Q1)
Coefficient of variation is defined as	(AM * 100)/S.D	(S.D* 100)/A.M	S.D/A.M	(1/S.D)*100	(S.D* 100)/A.M
In a symmetrical distribution	A.M =G.M=H.M	A.M >H.M >G.M	H.M >G.M >A.M	A.M <H.M <G.M	A.M =G.M=H.M
If the values of median and mean are 72 and 78 respectively, then find the mode.	16	60	70	76	60
If variance is 64, then find S.D.	8	13	14	11	8
Find Mean for the following 3, 4, 5.	4.25	2.25	3	2.28	3
The coefficient of range	L-S /L+S	L+S /L-S	L-S	L+S	L-S /L+S
Second quartile is also called as	Mode	mean	median	G.M	median
If S.D = 6, then find variance.	6	36	42	12	36
The mean of age of 5 men is 40 years. Three of them are of some age and they are excluded. The mean of the remaining two is 25. Age of one of the excluded person in years is:	20	25	40	50	50
If the mean of 50 observations is 50 and one observation 94 is wrongly recorded there as 49 then correct mean will be	49.1	50	50.9	58	50.9
Median is	Average point	Midpoint	Most likely point	Most remote point	Midpoint
Mode is the value which	Is a mid point	Occur the most	Average of all	Most remote Likely	Occur the most
. ……………. Is known as positional average	Median	Mean	Mode	Range	Median

Question	Option-1	Option-2	Option-3	Option-4	Answer
The median of marks 55, 60, 50, 40, 57, 45, 58, 65, 57, 48 of 10 students is	55	57	52.5	56	56
In a group of 150 observations the arithmetic mean is 60 and arithmetic mean of first 100 observations of the group is 50. Then arithmetic mean of the remaining observations of the group is	80	60	50	70	80
The middle most value of a frequency distribution table is known as	Mean	Median	Mode	Range.	Median
Quartiles are values dividing a given set of data into….. equal parts	4	6	3	2	4
The median value for the series 3, 5, 5, 2, 6, 2, 9, 5, 8, 6 is …	6	5	5.5	6.5	5
The mode for the series 3, 5, 6, 2, 6, 2, 9, 5, 8, 6 is	5	6	5.5	6.5	6
The Arithmetic mean for the series 51.6, 50.3, 48.9, 48.7, 48.5 is……….	49.8	50	48.9	49.6	49.8
The Median for the series 51.6, 50.3, 48.9, 48.7, 49.5, is…………	49.8	50	48.9	49.6	49.6
The Arithmetic mean for the series 51.6, 50.3, 48.9, 48.7, 49.5 is…………..	49.8	50	48.9	49.6	48.9
The Mode for the series 51.6, 50.3, 48.9, 48.7, 49.5 is…………	49.8	50	48.9	49.6	48.9
Mathematical ……………..is a positional average	Mean	median	mode	Standard deviation	median
The sum of deviations taken from arithmetic mean is ……………	minimum	zero	maximum	one	zero
The value of the variable which occurs most frequently in a distribution is called ….	Mean	median	mode	Standard deviation	mode
The formula of bimodal series is ……………..	Mode=2Median-3Mean	Mode= 3Median-2Mean	Mode= Median-Mean	Mode= Median-2Mean	Mode=3Median-2Mean
Deciles are the values dividing a given set of observations into	10	5	6	4	10
Percentiles divides a set of observations into	100	80	60	10	100
The middle most value of a frequency distribution table is known as	Mean	Median	Mode	Range	Median
Which of the following measures of averages divide the observation into two parts	Mean	Median	Mode	Range	Median
Which of the following measures of averages divide the observation into four equal parts	Mean	Median	Mode	Quartile	Quartile
The first quarter is known as	Lower quarter	Middle quarter	Upper quarter	Median	Lower quarter
The third quarter is known as	Lower quarter	Middle quarter	Upper quarter	Mode	Upper quarter
Arithmetic mean of the series 1, 3, 5, 7, 9 is	5	6	5.5	6.5	5
Arithmetic mean of the series 3, 4, 5, 6, 7 is ………	5.5	6	5	6.5	5
The Arithmetic mean for the series 3, 5, 5, 2, 6, 2, 9, 5, 8, 6, is…………….	5	6	5.5	6.5	5
Extreme values in a series affects the …………………….	Mean	median	mode	Standard deviation	Mean
Dispersion is also known as………..	Scatter	not scatter	line	nor line	Scatter

Question	Option-1	Option-2	Option-3	Option-4	Answer
The simple Range is …………..	R=L*S	R=L+S	R=L/S	R=L-S	R=L-S
The coefficient of variation is …………	100x	100x	Mean x		100x
Variance cannot be …………..	positive	negative	zero	one	negative
If A.M = 8, N=12, then find ∑X.	76	80	86	96	96
If the value of mode and mean is 60 and 66 then, find the value of median.	64	46	54	44	64
The formula for median for continuous series is	M = (N+1) / 2	M = L + [(N/2 + cf) / f] * i	M =L - (N/2+cf)/f* i	M = L + [(N/2 - cf) / f] * i	M = L + [(N/2 - cf) / f] * i
The formula for Q_1 for continuous series is	L + [(N/4 - cf) / f] * i	L +(N/2+cf)/f* I	L - (N/2-cf)/f*i	L + [(N/2 + cf) / f] * i	L + [(N/4 - cf) / f] * i
The formula for Q_3 for continuous series is	L +(N/2-cf)/f* i	L + [(3N/2 - cf) / f] * i	L - (3N/2-cf)/f*i	N/2 - cf	L + [(3N/2 - cf) / f] * i
If standard deviation is 5, then the variance is	5	625	25	2.23068	25
Standard deviation is also called as	Root mean square deviation	mean square deviation	Root deviation	Root median square deviation	Root mean square deviation
Measures of central tendency is also known as	Dispersion	averages	correlation	tendency	correlation
Q_1 = 40, Q_3 = 60 then coefficient of Q.D is	0.3	0.4	0.2	0.1	0.4
From the given data 35,40,43,32,27 the coefficient of range is	23	0.23	13	0.13	13
Sum of square of the deviations about mean is	Maximum	one	zero	Minimum	Minimum
Median is the value of ------------- item when all the items are in order of magnitude.	First	second	Middle most	last	Middle most
Find the Median of the following data 160, 180, 175, 179, 164, 178, 171, 164, 176.	160	175	176	180	175
The position of the median for an individual series is taken as	(N + 1) / 2	(N + 2) / 2	N/2	N/4	(N + 1) / 2
Mode is the value, which has	Average frequency density	less frequency density	greatest frequency density	graetest frequency	greatest frequency density
A frequency distribution having two modes is said to be	unimodal	bimodal	trimodal	modal	bimodal
Mode has --------------- stable than mean.	less	more	same	most	less
Which of the following is not a measure of dispersion?	Range	quartile deviation	standard deviation	median	median
Which one of the following shows the relation between variance and standard deviation?	var = square root of S.D	S.D = square root of variance	variance = S.D	variance / S.D = 1	S.D = square root of variance
Range of the given values is given by	L- S	L+S	S+L	LS	L- S

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Po),
Coimbatore –641 021
DEPARTMENT OF MATHEMATICS
PART-A Multiple Choice Questions (Each Question Carries One Mark)

Subject Name: Statistics for Business Decisions Subject Code: 17BAU102

UNIT-II

Question	Option-1	Option-2	Option-3	Option-4	Answer
Coefficient of correlation value lies between	1 and –1	0 and 1	0 and ∞	0 and –1.	1 and –1
While drawing a scatter diagram if all points appear to form a straight line getting Downward from left to right, then it is inferred that there is	Perfect positive correlation	simple positive correlation	Perfect negative correlation	no correlation	Perfect negative correlation
The range of the rank correlation coefficient is	0 to 1	–1 to 1	0 to ∞	– ∞ to ∞	–1 to 1
If r =1, then the angle between two lines of regression is	Zero degree	sixty degree	ninety degree	thirty degree	ninety degree
Regression coefficient is independent of	Origin	scale	both origin and scale	neither origin nor scale.	Origin
If the correlation coefficient between two variables X and Y is negative, then the Regression coefficient of Y on X is	Positive	negative	not certain	zero	negative
If the correlation coefficient between two variables X and Y is positive, then the Regression coefficient of X on Y is	Positive	negative	not certain	zero	Positive
There will be only one regression line in case of two variables if	r =0	r = +1	r = –1	r is either +1 or –1	r =0
The regression line cut each other at the point of	Average of X only	Average of Y only	Average of X and Y	the median of X on Y	Average of X and Y
If b_{xy} and b_{yx} represent regression coefficients and if $b_{yx} > 1$ then b_{xy} is	Less than one	greater than one	equal to one	equal to zero	Less than one
Which one of the following refers the term Correlation?	Relationship between two values	Relationship between two variables	Average relationship between two variables	Relationship between two things	Relationship between two variables
If r = +1, then the relationship between the given two variables is	perfectly positive	perfectly negative	no correlation	high positive	perfectly positive
If r = - 1, then the relationship between the given two variables is	perfectly positive	perfectly negative	no correlation	low Positive	perfectly negative
If r = 0, then the relationship between the given two variables is	Perfectly positive	perfectly negative	no correlation	both positive and negative	no correlation
If x and y are independent variables then,	cov(x,y)≠ 0	cov(x,y)= 1	cov(x,y)= 0	cov(x,y) >1	cov(x,y)= 0
Correlation coefficient is the ----------- of the two regression coefficients.	Mode	Geometric mean	Arithmetic mean	median	Geometric mean
b_{xy} = 0.4, b_{yx} = 0.9 then r =	0.6	0.3	0.1	-0.6	0.6
b_{xy}=1/5, r=8/15, s_x = 5 then s_y=	40/13	13/40	40/3	3	40/3
The geometric mean of the two regression coefficients.	Correlation coefficient	regression coefficients	coefficient of range	coefficient of variation	Correlation coefficient
If two variables are uncorrelated, then the lines of regression	Do not exist	coincide	Parallel to each other	perpendicular to each other	perpendicular to each other
If the given two variables are correlated perfectly negative, then	r = +1	r = -1	r = 0	r ≠ +1	r = -1

Question	Option-1	Option-2	Option-3	Option-4	Answer
If the given two variables have no correlation, then	$r = +1$	$r = -1$	$r = 0$	$r \neq +1$	$r = 0$
If the correlation coefficient between two variables X and Y is ---------, the Regression coefficient of Y on X is positive	Negative	positive	not certain	zero	positive
If the correlation coefficient between two variables X and Y is ---------, the Regression coefficient of Y on X is negative	Negative	positive	not certain	zero	Negative
The regression line cut each other at the point of----------	Average of X only	Average of Y only	Average of X and Y	the median of X on Y	Average of X and Y
Given the coefficient of correlation being 0.8, the coefficient of determination will be	0.98	0.64	0.66	0.54	0.64
Given the coefficient of correlation being 0.9, the coefficient of determination will be	0.98	0.81	0.66	0.54	0.81
If the coefficient of determination being 0.49, what is the coefficient of correlation	0.7	0.8	0.9	0.6	0.7
Given the coefficient of determination being 0.36, the coefficient of correlation will be	0.3	0.4	0.6	0.5	0.6
Maximum value of correlation is	2	1.5	1	0	1
Correlation between income and demand is	Negative	positive	zero	none of the above	positive
Minimum value of correlation is	-2	1.5	1	0	-1
Which is a method of measuring correlation?	Graphic correlation	scatter diagrams	both Graphic correlation and scatter diagrams	Either graphic correlation or scatter diagrams	both Graphic correlation and scatter diagrams
If there exists any relation between the sets of variables, it is called	regression	skewness	correlation	quartile	correlation
Rank correlation was discovered by	R.A.Fisher	Sir Francis Galton	Karl Pearson	Spearman	Spearman
Formula for Rank correlation is	$1- (6\Sigma d^2 /(n(n2-1)))$	$1- (6\Sigma d^2 /(n(n2+1)))$	$1+ (6\Sigma d^2 /(n(n2+1)))$	$1 /(n(n2-1))$	$1- (6\Sigma d^2 /(n(n2-1)))$
With $b_{xy}=0.5$, $r = 0.8$ and the variance of $Y=16$, the standard deviation of X=	6.4	2.5	10	25.6	2.5
The coefficient of correlation r =	$(b_{xy.} b_{yx})^{1/4}$	$(b_{xy.} b_{yx})^{-1/2}$	$(b_{xy.} b_{yx})^{1/3}$	$(b_{xy.} b_{yx})^{1/2}$	$(b_{xy.} b_{yx})^{1/2}$
If two regression coefficients are positive then the coefficient of correlation must be	Zero	negative	positive	one	positive
If two-regression coefficients are negative then the coefficient of correlation must be	Positive	negative	zero	one	Positive
The regression equation of X on Y is	$X = a + bY$	$X = a + bX$	$X= a - bY$	$Y= a + bX$	$X = a + bY$
The regression equation of Y on X is	$X = a + bY$	$X = a + bX$	$X= a - bY$	$Y= a + bX$	$Y= a + bX$
The given two variables are perfectly positive, if	$r = +1$	$r = -1$	$r = 0$	$r \neq +1$	$r = +1$
The relationship between two variables by plotting the values on a chart, known as-	coefficient of correlation	Scatter diagram	Correlogram	rank correlation	Scatter diagram
------------ is independent of origin and scale.	Correlation coefficient	regression coefficients	coefficient of range	coefficient of variation	Correlation coefficient
The angle between two lines of regression is ninety degree, if -------------	$r = 2$	$r = 0$	$r = 1$	$r = -1$	$r = 1$
--------- is used to measure closeness of relationship between variables.	Regression	mean	Rank correlation	correlation	correlation
If r is either +1 or –1, then there will be only one -------- line in case of two variables	Correlation	regression	rank correlation	mean	regression

Question	Option-1	Option-2	Option-3	Option-4	Answer
When b_{xy}=0.85 and b_{yx}= 0.89, then correlation coefficient r =	0.98	0.5	0.68	0.87	0.87
If b_{xy} and b_{yx} represent regression coefficients and if $b_{xy} < 1$, then b_{yx} is	less than 1	greater than one	equal to one	equal to zero	greater than one
While drawing a scatter diagram if all points appear to form a straight line getting Downward from left to right, then it is inferred that there is--------	Perfect positive correlation	simple positive correlation	Perfect negative correlation	no correlation	Perfect negative correlation
If r =1, the angle between two lines of regression is----------	Zero degree	sixty degree	ninety degree	thirty degree	ninety degree
Regression coefficient is independent of--------	Origin	scale	both origin and scale	neither origin nor scale.	Origin
There will be only one regression line in case of two variables if------------	r =0	r = +1	r = –1	r is either +1 or –1	r =0
Which of the following measurement scales is required for the valid calculation of Karl Pearson's correlation coefficient?	ordial	interval	ratio	nomial	interval
which of the following is the highest range of r?	0 and 1	-1 and 1	-1 and 0	1 and 2	-1 and 1
Given the following details, find the value of $\sigma_{Y,}$ r = 0.6, Cov(X,Y) = 12, σ_{X-} = 5	4	5	6	9	4
What will be the range of r when we find that the dependent variable increases as the independent variable increases?	0 to -0.005	1 to 2	0.1 to 1	0.05 to 1	0.1 to 1
When the two regression lines coincide, then r is	0	-1	1	2	1
Which one of the following refers the term Correlation?	0	-1	1	2	2

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Po),
Coimbatore –641 021
DEPARTMENT OF MATHEMATICS
PART-A Multiple Choice Questions (Each Question Carries One Mark)

Subject Name: Statistics for Business Decisions **Subject Code: 17BAU102**

UNIT-III

Question	Option-1	Option-2	Option-3	Option-4	Answer
A ------- is an arrangement of statistical data in a chronological order.	time series	data	forecasting	index number	time series
Time series helps in ----------.	forecasting	evaluation	comparison	all the above	all the above
There are ------- types of components of a time series.	3	4	2	5	4
The ------------ model assumes that the observed value is the sum of four component of time series	multiplicative	secular	additive	cyclical	additive
The ---------- model assumes that the observed value is obtained by multiplying the trend by the rates of three other components	multiplicative	secular	additive	cyclical	multiplicative
Seasonal variations repeat during a period of----- years.	5	1	7	3	1
The most important factor causing seasonal variations is-----	depression in business	growth of population	weather and social customs	none of these	weather and social customs
If the trend is absent, the seasonal indices are known by--------	ratio to trend method	simple average method	ratio to moving average method	none of these	simple average method
The trend can be found by the method of least squares if the --------	trend is not clear	the trend is linear	trend is not linear	none of these	trend is not linear
The trend is linear if ---------	rate of growth is positive	growth rate is constant	growth is not constant	none of these\	growth rate is constant
The most widely used method of measuring seasonal variations is ---------	ratio to trend method	link relative	ratio to moving average	none of these	ratio to moving average
The ----------------may be used either to fit a straight line trend or a parabolic trend.	graphic method	method of least squares	semi average method	moving average method	method of least squares
Whenever we fit any straight line trend by the least squares method,which things should be specified?	which year was selected as the origin?	what is the unit of time represented by X?	In what kind of units is Y being measured?	all the above	all the above
The simplest example of the --------- is the second degree parabola.	linear trend	secular trend	non-linear trend	none of the above	non-linear trend
In second deree parabola when time origin is taken between two middle years ∑X would be --------.	1	2	3	0	0
Trends may also be plotted on a semi-loarithmic chart in the form of a-------	straight line	non-linear curve	either a or b	none of these	either a or b
How many types of trend are usually computed by logarithms?	1	2	3	4	2
The types of trend usually computed by logarithms are ----------	exponential tends	growth curves	both a and b	none of these	both a and b
A ----------- variations repeat during a period of 1 year.	seasonal	secular	cyclical	irregular	seasonal

Question	Option-1	Option-2	Option-3	Option-4	Answer
The------ helps in forescating,evaluation and comparison.	time series	data	correlation	index number	time series
The trend is ----------- if growth rate is constant.	non-linear	linear	clear	none of these	linear
The most important factor causing -------- variations is weather and social customs.	seasonal	secular	cyclical	irregular	seasonal
The simplest example of the non-linear trend is the second degree --------	ellipse	parabola	hyperbola	circle	parabola

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Po),
Coimbatore –641 021
DEPARTMENT OF MATHEMATICS
PART-A Multiple Choice Questions (Each Question Carries One Mark)

Subject Name: Statistics for Business Decisions **Subject Code: 17BAU102**

UNIT-IV

Question	Option-1	Option-2	Option-3	Option-4	Answer
Index numbers are special type of ----------	averages	percentages	economic activity	time series	time series
The base period should not be too distant from the -----	current	past	future	arbitrary	current
A good index number is one that satisfies----------------------------	unit test	time reversal test	factor test and circular test	all the above	all the above
-------------- help to calculate the real wages	Index numbers	averages	time series	trend	Index numbers
The best average to calculate index number is -------------	A.M	G.M	H.M	both A.M and G.M	G.M
Current year quantity is used in -------------	simple A.M method	Kelly's method	Laspeyre's method	Fisher's method	Fisher's method
Laspeyre's index is based on-------------------------	Base year quartiles	Current year quartiles	Both of them	Average of current and base year	Base year quartiles
Time Reversal test is satisfied by------------------------------	Laspeyre's method	Paasche's method	Fisher's method	Bow ley's method	Fisher's method
If the price of a commodity is Rs.80 in the base year and Rs.72 in the current year, the Price index number is-------	8	90	111.11	110	90
In Laspeyre's index number, importance is given to the quantity of---------	current year	base year	future year	arbitrary year	base year
The current year quantities are taken as weights in------	Bow ley's formula	Laspeyre's formula	Paasche's formula	Fisher's formula	Paasche's formula
Time reversal test condition is--------	P01 xP10 = 0	P_{01} xP_{10} < 1	P_{01} xP_{10} = 1	P_{01} xP_{10} > 1	P_{01} xP_{10} = 1
Paasche's index number is generally expected to have--------	an upward bias	a downward bias	either upward or downward bias	no bias	an upward bias
The formula for unweighted averages of relatives' method by using A.M.,	å log P/N	åP/N	N/åP	åP/ log N	åP/N
Fisher's ideal index is--------	arithmetic mean of Laspeyre's and Paasche's index	geometric mean of Laspeyre's and Paasche's index	median of Lasoeyre's and Paasche's index	all of the above	geometric mean of Laspeyre's and Paasche's index
If åp_0 is 3100 and åp_1 is 4700 then P_{01} =--------------	151	150	151.61	125.2	151.61
P = √P_{01}^{L} x P_{01}^{P} is -------------------	Laspeyre's formula	Paasche's formula	Fisher's formula	both a) and b)	Fisher's formula
Family budget method is a method to calculate ----- price index.	consumer	Laspeyre's	Paasche's	Fisher	consumer
The best average in the construction of index number is	median	geometric mean	arithmetic mean	mode	geometric mean
Commodities which show considerable price fluctuation could be best measured by a	value index	price index	quantity index	quality index	quantity index
The circular test is an extension of the	factor reversal test	time reversal test	t-test	f-test	time reversal test
Most frequently used index number formulae are	weighted	fixed weighted	un weighted	fixed un weighted	weighted
Laspeyre's index is based on	base year quantities	current year quantities	base year qualities	current year qualities	base year quantities

Question	Option-1	Option-2	Option-3	Option-4	Answer
------- index number uses the geometric mean of the base year and current year quantities as weights.	Kelly's	Walsch's	Fisher's price	Marshall-Edgeworth's	Walsch's
----- is the sum of the values of a given year divided by the sum of the values of the base year.	Value index	Laspeayre's index	Paasche's index	Fisher ideal index	Value index
Formula for price relative or price index number of a commodity P is-----	$(p_1/p_0)*100$	$(p_1/q_0)*100$	$(q_1/p_1)*100$	$(p_1/q_1)*100$	$(p_1/p_0)*100$
Fisher's formula is called --------- index number formula	Ideal	economic	special	commercial	Ideal
Factor reversal test is satisfied by------------	Laspeyre's method	Paasche's method	Fisher's method	both a) and b)	Fisher's method
The year for which index number is calculated is called ---------year.	current	base	average	calculated	current
Notation of price of a commodity in the current year is-------	p_1	p_0	q_0	q_1	p_1
---------- are the pulse of an economy	Time series	Mean	Mode	Index number	Index number
Purchasing power = ------------	100/Price index	Price index /100	Money wage/Price index *100	Price index *100	100/Price index
Notation of price of a commodity in the base year is-------	p_1	p_0	q_0	q_1	p_0
Notation of quantity of a commodity in the current year is-------	p_1	p_0	q_0	q_1	q_1
If the price of a commodity is Rs.40 in the base year and Rs.50 in the current year, the Price has increased by-------	25%	10%	125%	35%	25%
By circular test ---------------	$P_{01} \times P_{12} \times P_{20} = 1$	$P_{12} \times P_{20} = 1$	$P_{01} \times P_{12} = 1$	$P_{01} \times P_{12} \times P_{20} = 0$	$P_{01} \times P_{12} \times P_{20} = 1$
Link relative is a price or quantity relative with the condition that ------------- is the preceding year.	current year	base year	arbitrary year	previous year	base year
Cost of living index number is also known as -----------	Consumer price	Consumer price index number	Consumer price index number	price index number	Consumer price index number
In Factor reversal test P_{01} gives the relative change in -----------	price	quantity	both	neither price nor quantity	price
In Factor reversal test Q_{01} gives the relative change in -----------	price	quantity	both	neither price nor quantity	quantity
----------- satisfies the Kelly's test.	Time reversal test	Factor reversal test	Fisher's test	both a and b	Time reversal test
Fisher's index satisfy ------------	Time reversal test	Factor reversal test	Fisher's test	both a and b	both a and b
In Factor reversal test $P_{01} \times Q_{01}$ gives the relative change in -----------	price multiplied by quantity	quantity	both	price	price multiplied by quantity
Index number are expressed in ------	ratio	percentage	fraction	mean	percentage
Paasche index is based on	base year quantities	current year quantities	base year qualities	current year qualities	current year quantities
The ------------- are special type of time series.	forecasting	evaluation	comparison	index numbers	Index numbers

KARPAGAM ACADEMY OF HIGHER EDUCATION
(Deemed to be University Established Under Section 3 of UGC Act 1956)
Pollachi Main Road, Eachanari (Po),
Coimbatore –641 021
DEPARTMENT OF MATHEMATICS
PART-A Multiple Choice Questions (Each Question Carries One Mark)

Subject Name: Statistics for Business Decisions **Subject Code: 17BAU102**

UNIT-V

Question	Option-1	Option-2	Option-3	Option-4	Answer
If binomial distribution is symmetrical if p=q=?	1	0.4	0	0.5	0.5
Binomial distribution is positively skewed if	p>0.5	p<0.5	p=0.5	p = 0	p<0.5
Binomial distribution is negatively skewed if	p = 0	p<0.5	p=0.5	p>0.5	p>0.5
The probability of drawing diamond and a heart card from a pack of 52 cards is	13/102	(1/4)	2/13	(7/16)	13/102
The probability of drawing king and queen card from a pack of 52 cards is	13/102	(1/4)	(2/13)	8/663	8/663
The probability of drawing a card of King from a pack of cards is	(1/4)	1/11	1/12	1/13	1/13
In the case of poisson distribution , if the mean is 4, the standard deviation is ,	16	4	2	1	2
For a poisson distribution	mean < Variance	mean = Variance	mean > Variance	mean < Variance	mean = Variance
In coin, the probability of getting head is	(1/2)	(1/3)	2	0	(1/2)
The probability that a leap year selected at random contain 53 Sundays is	(1/7)	(2/7)	(3/7)	(1/53)	(2/7)
The probability of drawing king and queen card from a pack of 52 cards is	13/102	(1/4)	(2/13)	8/663	8/663
Two coins are tossed five times, find the probability of getting an even number of heads	0.25	1	0.4	0.25	0.25
Mean of a Binomial distribution is 24, Standard deviation = 4, n, p, q respectively are :	72, 1/3, 2/3	60, 1/3, 2/3	87, ¼, 3/4	90, 1/5, 4/5	72, 1/3, 2/3
1 out of 10 electrical switches inspected are likely to be defective. The mean and standard deviation of 900 electrical switches inspected are	90, 9	81, 9	88, 10	91, 11	81, 9
If the mean of a Poisson distribution’s 4, find S.D.	0.25	2	3.25	4	4
If the mean of a binomial distribution is 5 and standard deviation 2 find the number of items in the distribution	20	25	16	9	25
In a binomial distribution mean and mode are equal only when	P=0.5	p=0.9	q=0.1	all the situations	P=0.5
The variance of a binomial distribution is measured by	np	np(1 – p)	Pq	Nq	np(1 – p)
The mean of binomial distribution is measured by	np	npq	Pq	Nq	np
What is the chance of getting a king in a draw from a pack of 52 cards?	1/52	1/12	1/13	1/14	1/52
The probability of drawing a card of King from a pack of cards is	(1/4)	1/11	1/12	1/13	1/13
In the case of poisson distribution , if the mean is 4, the standard deviation is ,	16	4	2	1	2
For a poisson distribution	mean < Variance	mean = Variance	mean > Variance	mean < Variance	mean = Variance
In coin, the probability of getting head is	(1/2)	(1/3)	2	0	(1/2)

Question	Option-1	Option-2	Option-3	Option-4	Answer
The probability that a leap year selected at random contain 53 Sundays is	(1/7)	(2/7)	(3/7)	(1/53)	(2/7)
The square of the S.D is	variance	coefficient of variation	square of variance	square of coefficient of variation	variance
A bag contains 7 red and 8 black balls. The probability of drawing a red ball is	7/15	8/15	1/15	14/15	1/15
For Binomial distribution ,mean is	npq	n	p	np	np
The probability of drawing a card of clubs from a pack of 52 cards is	(1/52)	(1/3)	¼	1/13	1/13
The probability of drawing an ace or queen card from a pack of 52 cards is	1/13	1/4	2/13	1/52	1/13
The total probability is	0.5	2	1	0	1
Two cards are drawn from a pack of 52 cards. Find the probability that both are red cards.	26C2	52C4	52C2	26C3	26C2
Which of the following is true for a Poisson distribution	Mean > Variance	Mean < Variance	Mean = Variance	Mean = SD	Mean = Variance
The mean of Binomial distribution is measured by	np	npq	pq	nq.	np
Two coins are tossed simultaneously. What is the probability of getting a head and a tail?	1/4	4/4	2/4	¾	1/4
Which of the following is true for a binomial distribution	Mean > Variance	Mean < Variance	Mean = Variance	Mean = SD	Mean < Variance
One card is drawn at random from a well-shuffled pack of 52 cards. What is the probability that it will be a diamond ?	1/13	1/4	1/52	1/15	1/13
The square of the S.D is	variance	coefficient of variation	square of variance	square of coefficient of variation	variance