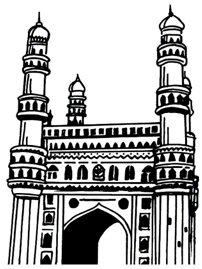
Rahul's ✔

Topper's Voice



B.Com.

II Year IV Semester

Latest 2021-22 Edition

BUSINESS STATISTICS-II

- **Study Manual**
- **FAQ's and Important Questions**
- **Short Question & Answers**
- **Appendix**
- **Choose the Correct Answers**
- **Fill in the blanks**
- **Exercise Problems**
- **Solved Previous Question Papers**
- **Solved Model Papers**

` .179/-

- by -

WELL EXPERIENCED LECTURER

Rahul Publications™

Hyderabad. Ph : 66550071, 9391018098

B.Com.

II Year IV Semester

BUSINESS STATISTICS-II

Inspite of many efforts taken to present this book without errors, some errors might have crept in. Therefore we do not take any legal responsibility for such errors and omissions. However, if they are brought to our notice, they will be corrected in the next edition.

Price **`**. *179/-*

Sole Distributors : ☎ : **66550071, Cell : 9391018098**

VASU BOOK CENTRE

Shop No. 3, Beside Gokul Chat, Koti, Hyderabad.

Maternity Hospital Opp. Lane, Narayan Naik Complex, Koti, Hyderabad.
Near Andhra Bank, Subway, Sultan Bazar, Koti, Hyderabad -195.

CONTENTS

BUSINESS STATISTICS-II

STUDY MANUAL

SOLVED PREVIOUS QUESTION PAPERS

SOLVED MODEL PAPERS

SYLLABUS

UNIT - I

REGRESSION :

Introduction - Linear and Non Linear Regression - Correlation Vs. Regression - Lines of Regression - Derivation of Line of Regression of Y on X - Line of Regression of X on Y - Using Regression Lines for Prediction.

UNIT - II

INDEX NUMBERS :

Introduction - Uses - Types - Problems in the Construction of Index Numbers - Methods of Constructing Index Numbers - Simple and Weighted Index Number (Laspeyre - Paasche, Marshall – Edgeworth) - Tests of Consistency of Index Number: Unit Test - Time Reversal Test - Factor Reversal Test - Circular Test - Base Shifting - Splicing and Deflating of Index Numbers.

UNIT - III

TIME SERIES :

Introduction - Components - Methods-Semi Averages - Moving Averages - Least Square Method - Deseasonalisation of Data - Uses and Limitations of Time Series.

UNIT - IV

PROBABILITY :

Probability - Meaning - Experiment - Event - Mutually Exclusive Events - Collectively Exhaustive Events - Independent Events - Simple and Compound Events - Basics of Set Theory - Permutation- Combination - Approaches to Probability: Classical - Empirical - Subjective - Axiomatic - Theorems of Probability: Addition - Multiplication - Baye's Theorem.

UNIT - V

THEORITECAL DISTRIBUTIONS:

Binomial Distribution: Importance - Conditions - Constants - Fitting of Binomial Distribution. Poisson Distribution: - Importance - Conditions - Constants - Fitting of Poisson Distribution. Normal Distribution: - Importance - Central Limit Theorem - Characteristics - Fitting a Normal Distribution (Areas Method Only).

Contents

UNIT - I

UNIT - II

Frequently Asked and Important Questions

UNIT - I

1. What is the importance of regression analysis ?

Ans : **(Jan.-21, Imp.)**

Refer Unit-I, Q.No. 2

2. Explain different types of Regression.

Ans : **(Imp.)**

Refer Unit-I, Q.No. 4

3. What are the limitations of regression analysis ?

Ans : **(Jan.-21, Imp.)**

Refer Unit-I, Q.No. 5

4. Differentiate between linear and non-linear regression.

Ans : **(Imp.)**

Refer Unit-I, Q.No. 7

5. Differentiate between Correlation and Regression.

Ans : **(June-18, Imp.)**

Refer Unit-I, Q.No. 8

6. What do you mean by line of regression? Derive the equations of lines of regression.

Ans : **(Imp.)**

Refer Unit-I, Q.No. 9

7. What are the disadvantages of regression analysis?

Ans : **(Jan.-21, Imp.)**

Refer Unit-I, Q.No. 12

8. Given :

$\Sigma x=56, \Sigma y=40, \Sigma x^2=524, \Sigma y^2=256, \Sigma xy=364, N = 8$

(i) Find the two Regression equations and

(ii) The Correlation Coefficient.

Sol: **(June-18, Imp.)**

Refer Unit-I, Prob. 8

9. Following are the marks in Statistics and English in an Annual Examination.

Particular	Statistics (X)	English (Y)
Mean	40	50
Standard Derivation	10	16
Co-efficient Correlation	0.5	

(i) Estimate the score of English, when the score in Statistics is 50.

(ii) Estimate the score of statistics, when the score in English is 30.

Sol: **(June-19, Imp.)**

Refer Unit-I, Prob. 9

10. You are given the following information about advertisement expenditure and sales :

Particulars	Adv. Exp (X) (` Crores)	Sales (Y) (` Crores)
Mean	20	120
S.D	5	25
Correlation coefficient	0.8	

(i) Calculate two regression equations

(ii) Find likely sales when Adv. Expenses is ` 25 Crores

(iii) What should be the Adv. Budget if the company wants to attain sales target of ` 150 crores.

Sol : **(Imp.)**

Refer Unit-I, Prob. 10

UNIT - II

1. Define Index Numbers.

Ans : **(Jan.-21, June-19, Imp.)**

Refer Unit-II, Q.No. 1

2. What are the uses of index numbers ?

Ans : **(June-19, June-18, Imp.)**

Refer Unit-II, Q.No. 2

3. What are the characteristics of Index Numbers ?

Ans : **(June-19, Imp.)**

Refer Unit-II, Q.No. 3

4. Explain the various types of index numbers.

Ans : **(June-19, Imp.)**

Refer Unit-II, Q.No. 4

5. What are the problems involved in construction of index numbers ? Explain.

Ans : **(Imp.)**

Refer Unit-II, Q.No. 5

6. What are the various methods of Constructing Index Numbers?

Ans : **(Imp.)**

Refer Unit-II, Q.No. 6

7. Explain the various tests of Consistency of Index Number.

Ans : **(Imp.)**

Refer Unit-II, Q.No. 12

8. From the following data construct on Index Number for 2017 taking 2016 as base as per simple aggregative Method.

Commodity	P	Q	R	S	T
Price 2015 (`)	40	60	85	25	30
Price 2017 (`)	60	90	125	30	40

Sol : **(Jan.- 21, Imp.)**

Refer Unit-II, Prob. 4

9. From the following data calculate price index according to

(i) Laspeyre,

(ii) Paasche and

(iii) Marshall-Edgeworth methods.

Item	Base year		Current year	
	Price (`)	Expenditure (`)	Price (`)	Expenditure (`)
A	5	50	8	40
B	7	25	12	30
C	9	10	15	25
D	12	5	20	18

Sol : **(Jan.-21)**

Refer Unit-II, Prob. 11

10. From the following data calculate price index Number by using (i) Paasche's Method and (ii) Marshal Edgeworth Method.

Item	Base year		Current year	
	Price (Rs.)	Expenditure (Rs.)	Price (Rs.)	Expenditure (Rs.)
P	6	300	10	560
Q	2	200	2	240
R	4	240	6	360
S	10	300	12	288
T	3	120	8	240

Sol : **(June-18, Imp.)**

Refer Unit-II, Prob. 13

11. The following are the indices (2007. Base)

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Indices	100	120	122	116	120	120	137	136	149	156	137

Shift the base to 2012 and recast the index numbers.

Sol. **(June-18, Imp.)**

Refer Unit-II, Prob. 18

UNIT - III

1. Define time series.

Ans : **(Imp.)**

Refer Unit-III, Q.No.1

2. What are the characteristics of time series ?

Ans : **(Imp.)**

Refer Unit-III, Q.No. 2

3. What are the Components of Time Series?

Ans : **(Jan.-21, Imp.)**

Refer Unit-III, Q.No. 4

4. What are the various methods of time series? Explain in detail semi average method. With an example.

Ans : **(Imp.)**

Refer Unit-III, Q.No. 9

5. Discuss the method of moving averages in measuring trend. What are it merits and limitations of moving average method?

Ans : **(Imp.)**

Refer Unit-III, Q.No. 11

6. Define least square method. Explain merits and demerits of least square method.

Ans : **(Imp.)**

Refer Unit-III, Q.No. 12

7. Explain the uses and limitations of time series.

Ans : **(June-19, June-18, Imp.)**

Refer Unit-III, Q.No. 14

8. From the following data fit a trend line by the method of Semi-Average.

Year :	2012	2013	2014	2015	2016	2017
Output :	20	16	24	30	28	32

Sol : **(Jan.-21, Imp.)**

Refer Unit-III, Prob. 10

9. Find the 4 yearly moving averages from the following data.

Year	2008	209	2010	2011	2012	2013	2014	2015	2016	2017
Production (in Tonnes)	150	170	196	180	190	216	248	280	300	320

Sol : **(June-18, Imp.)**

Refer Unit-III, Prob. 14

10. Calculate the 3 yearly the 5 yearly moving averages for the following time series.

Year :	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Production (in quintals)	500	540	550	530	520	560	600	640	620	610	640

Sol : **(Jan.-21, Imp.)**

Refer Unit-III, Prob. 16

11. Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

Sol : **(June-19, Imp.)**

Refer Unit-III, Prob. 18

UNIT - IV

1. Explain the various terms used in probability theory.

Ans : **(Jan.-21, June-18, Imp.)**

Refer Unit-IV, Q.No. 3

2. Explain in detail abiout Permutation and Combination.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 5

3. Explain the Axiomatic approach to probability.

Ans : **(Jan.-21, Imp.)**

Refer Unit-IV, Q.No. 10

4. State and explain Baye's probability theorem.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 13

5. What are the applications of Baye's Theorem?

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 14

6. A bag contains 4 defective and 6 good Electronic Calculators. Two calculators are drawn at random one after the other without replacement. Find the probability that

i) Two are good

ii) Two are defective and

iii) One is good and one is defective.

Sol : **(June-19, Imp.)**

Refer Unit-IV, Prob. 19

7. One card is drawn at random from a pack of 52 cards. What is the probability that it is either a king or a queen?

Sol : **(June-19, Imp.)**

Refer Unit-IV, Prob. 18

8. A box contains 8 Red and 5 White balls. Two successive draws of 3 balls are made at random. Find the probability that the first three are white and second three are red.

(i) When there is replacement and

(ii) When there is no replacement.

Sol : **(Jan.-21, Imp.)**

Refer Unit-IV, Prob. 26

9. From 30 tickets marked with first 30 numerals, 1 ticket is drawn at random. It is then replaced and a second draw is made. Find the probability that in the first draw it is multiple of 5 or 7 and in second draw it is a multiple of 3 or 7.

Sol : **(June-18, Imp.)**

Refer Unit-IV, Prob. 27

---

10. A factory has two machines. Empirical evidence has established that machines I and II produce 30% and 70% of output respectively. It has also been established that 5% and 1% of the output produced by these machines respectively was defective. A defective item is drawn at random. What is the probability that the defective item was produced by either machine I or machine II ?

Sol : **(Jan.-21, Imp.)**

Refer Unit-IV, Prob. 30

---

11. A company has two plants for manufacturing scooters. Plant I manufacturers 80% of the Scooters and Plant II manufactures 20%. At the Plant I 85% Scooters are rated to be of standard quality and at plant II 65% Scooters are rated to be of standard quality. One Scooter was selected at random. What is the probability that

i) It is manufactured by Plant I

ii) It is manufactured by Plant II – which is of standard quality.

Sol : **(June-19, Imp.)**

Refer Unit-IV, Prob. 32

UNIT - V

1. What is binomial distribution? State the importance of binomial distribution.

Ans : **(Imp.)**

Refer Unit-V, Q.No.1

---

2. A die is tossed thrice. A success is getting 1 or 6 on a toss. Find the mean and variance of the number of successes.

Sol : **(June-19, Imp.)**

Refer Unit-V, Prob. 7

---

3. 8 Coins are tossed at a time, 256 times. Find the expected frequencies of successes Getting a Head) and tabulate the results obtained.

Sol : **(Jan.-21, Imp.)**

Refer Unit-V, Prob. 8

4. Ten unbiased coins are tossed simultaneously. Find the probability of obtaining:

i) Exactly 6 Heads

ii) Atleast 8 Heads

iii) No Heads

iv) Atleast one Head

v) Not more than 3 Heads and

vi) Atleast 4 heads.

Sol : **(June-19, Imp.)**

Refer Unit-V, Prob. 10

5. What are the features and assumptions of Poisson distribution?

Ans : **(Imp.)**

Refer Unit-V, Q.No. 7

6. What are the properties of normal distribution?

Ans : **(June-18, Imp.)**

Refer Unit-V, Q.No. 15

7. Explain in detail about Central Limit Theorem.

Ans : **(Imp.)**

Refer Unit-V, Q.No. 16

UNIT I	**REGRESSION :** Introduction - Linear and Non Linear Regression - Correlation Vs. Regression - Lines of Regression - Derivation of Line of Regression of Y on X - Line of Regression of X on Y - Using Regression Lines for Prediction.

1.1 Regression

1.1.1 Introduction

Q1. What do you understand by Regression?

(OR)

What is meant by Regression?

(OR)

Define Regression?

Ans : **(Jan.-21, June-18)**

Regression analysis which confines itself to a study of only two variables is called simple regression. The regression analysis which studies more than two variables at a time is called multiple regression. In the simple regression analysis there are two variables-one of which is known as 'independent variable' or 'regressor' or 'predictor'. On the basis of the values of this variable the values of the other variable are predicted. The other variable whose values are predicted is called the 'dependent' or 'regressed' variable.

Definitions

1. "Regression is the measure of the average relationship between two or more variables in terms of the original units of the data."
2. **According to Morris Hamburg** The term 'regression analysis' refers to the methods by which estimates are made of the values of a variable from a knowledge of the values of one or more other variables and to the measurement of the errors involved in this estimation process."
3. **According to Taro Yamane** "One of the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related causally, is regression analysis."
4. **According to YaLum Chou** "Regression analysis attempts to establish the 'nature of the relationship between variables that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting."

Q2. What is the importance of regression analysis ?

Ans : **(Jan.-21, Imp.)**

1. Regression analysis helps in establishing a functional relationship between two or more variables. Once this is established it can be used for various advanced analytical purposes.
2. Since most of the problems of economic analysis are based on cause and effect relationship, the regression analysis is a highly valuable tool in economics and business research.
3. This can be used for prediction or estimation of future production, prices, sales, investments, income, profits and population which are indispensable for efficient planning of an economy and are of paramount importance to a businessman or an economist.

4. Regression analysis is widely used in statistical estimation of demand curves, supply curves, production functions, cost functions, consumption functions, etc. Economists have discovered many types of production functions by fitting regression lines to input and output data.

Q3. What are the objectives of Regression Analysis.

Ans :

1. The first objective of regression analysis is to provide estimates of values of the dependent variable from values of independent variable. This is done with the help of the regression line. The regression line describes the average relationship existing between X and Y variables, more precisely, it is a line which displays mean values of Y for given values of X.

2. The second objective of regression analysis is to obtain a measure of the error involved in using the regression line as a basis for estimation. For this purpose standard error of estimate is obtained. This helps in understanding the correlation existing between X and Y.

3. In general, we can model the expected value of y as an n^{th} order polynomial, yielding the general polynomial regression model

$$Y = a_0 + a_1x + a_2x^2 + a_2x^2 + a_3x^3 + \dots + a_nx^n + \varepsilon$$

Conveniently, these models are all linear from the point of view of estimation, since the regression function is linear in terms of the unknown parameters ao,ai,.... Therefore, for least squares analysis, the computational and inferential problems of polynomial regression can be completely addressed using the techniques of multiple regressions. This is done by treating x, x^2, ... as being distinct independent variables in a multiple regression model.

Q4. Explain different types of Regression.

Ans : **(Imp.)**

The various types of Regression are as follows:

1. Simple Regression

In statistics, simple regression is the least squares estimator of a linear regression model with a single predictor variable. In other words, simple linear regression fits a straight line through the set of n points in such a way that makes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible.

2. Multiple Regression

Multiple regression analysis represents a logical extension of two-variable regression analysis. Instead of a single independent variable, two or more independent variables are used to estimate the values of a dependent variable. However, the fundamental concepts in the analysis remain the same.

3. Linear Regression

Linear regression is a form of regression which is used for modeling the relationship between scalar variables like X and F under linear regression, linear functions are used to model the data and the unknown parameters, of models are estimated from the data. Hence, these models are known as linear models.

4. Non-linear regression

In the non-linear regression the explained variable (dependent variable) changes at varying rate with a given change in the explaining variable (independent variable). It is also known as Curvilinear regression. Under non linear regression, the observational data are modeled by a function i.e., a non linear blend of model parameters and depends on one or more independent variable.

Q5. What are the limitations of regression analysis ?

Ans : **(Jan.-21, Imp.)**

Limitations of Regression Analysis

1. It assumes a linear relationship between two variables which need not be the case always.
2. It assumes a static relationship between the two variables over a period of time. However, relationships between variables can change with a change in other factors. For example, the change in demand for a given change in price can be estimated using regression. However, the impact of price on demand will be different when a family or a nation is poor and when such a family or nation has abundance of wealth or resources.
3. Regression analysis provides meaningful insights only up to a certain limit. For example, increasing production results in a decrease in marginal cost. However, beyond a certain point, increase in production can result in the costs going up.

1.1.2 Linear and Non Linear Regression

Q6. What do you mean by Linear and Non Linear Regression?

Ans :

(i) Linear Regression

Linear regression is a form of regression which is used for modeling the relationship between scalar variables like X and F under linear regression, linear functions are used to model the data and the unknown parameters, of models are estimated from the data. Hence, these models are known as linear models.

Linear models more commonly refers to those models, where the conditional mean of variable 'F for a given value of variable X will be an affine function of X. A linear regression may also refer to a model, where median or other quantile of the conditional distribution of 'F for a given value of 'X is termed as linear function of X. Similar, to all types of regression analysis, linear regression also aims on the conditional probability distribution of ' F for a given 'X, instead of joint probability distribution of 'F and X.

(ii) Non-Linear Regression

In the non-linear regression the explained variable (dependent variable) changes at varying rate with a given change in the explaining variable (independent variable). It is also known as Curvilinear regression. Under non linear regression, the observational data are modeled by a function i.e., a non linear blend of model parameters and depends on one or more independent variable.

Method of successive approximations are used for fitting the data. The data in non linear regression contains of error free independent variable 'X' and its relatively observed dependent variable 'Y.

Example

The output of rice increases rapidly with the application of the initial dose of fertilizer; there after it increases at a falling rate. The relationship in such case, when shown on graph will yield a 'curve'.

Q7. Differentiate between linear and non-linear regression.

(OR)

Compare and contrast linear and non-linear regression.

Ans : **(Imp.)**

The differences between linear and non-linear regression are as follows,

S.No.	Basis	Linear Regression	Non-Linear Regression
1.	Meaning	Linear regression is a form of regression which is used for modelling the relation ship between a scalar variable 'X' and 'Y'.	Non linear regression is a type of regression, where the observational data are modeled by a function i.e., a non linear blend of model parameters.
2.	Curve	If the regression curve is a straight line, then the regression is termed as linear regression.	If the curve of the regression is not a straight line. then the regression is termed as curved or non-linear regression.
3.	Model form	Under this, the parameters are considered as linear combinations.	Under this, the parameter are considered as functions
4.	Solution	Under linear regression, the solution for parameters is represented as closed form	Under non linear regression it is necessary for parameters to be solved repeatedly by using optimization algorithms.
5.	Uniqueness	The solution under linear regression is unique.	The Sum of the Squared Errors (SSE) may not be appear as unique.
6.	Parameters estimation	In case of uncorrelated errors, estimation Parameters are unbiased.	Incase of uncorrelated errors, estimation of Parameters are usually biased.
7.	Equation	The equation of regression curve is the equation of a straight line i.e., first degree equation in variables X and Y.	The regression equation will be functional relation between variables X and Y involving terms in x and Y involving terms in x and y of degree more than one.

1.2 Correlation Vs. Regression

Q8. Differentiate between Correlation and Regression.

(OR)

What are the differences between Correlation and Regression.

Ans : **(June-18, Imp.)**

S.No.	Basis for Comparison	Correlation	Regression
1.	Meaning	Correlation is a statistical measure which determines co-relationship or association of two variables.	Regression describe how an independent variable is numerically related to the dependent variable.
2.	Usage	To represent linear relationship between two variables.	To fit a best line and estimate one variable on the basis of another variable.
3.	Dependent and Independent variables	No difference	Both variables are different
4.	Indicates	Correlation coefficient indicates the extent to which two variables move together.	Regression indicates the impact of a unit change in the known variable (x) on the estimated variable (Y).
5.	Objective	To find a numerical value expressing the relationship between variables.	To estimate values of random variable on the basis of the values of fixed variable.

1.3 Lines of Regression

1.3.1 Derivation of Line of Regression of Y on X - Line of Regression of X on Y

Q9. What do you mean by line of regression? Derive the equations of lines of regression.

Ans : **(Imp.)**

In a bi-variate distribution, if the variables are related then the points when plotted in the scatter diagram will lie near a straight line which is called the line of regression and the regression is said to be linear regression. If points lie on some non-linear curve then the regression is said to be curvilinear regression.

(I) Regression of Y on X.

The regression equation of Y on X is expressed as follows :

Y = a+bX

It may be noted that in this equation 'Y' is a dependent variable, i.e., its value depends on X. 'X is independent variable, i.e., we can take a given value of X and compute the value of Y.

'a' is "Y-intercept" because its value is the point at which the regression line crosses the Y-axis, that is, the vertical axis, 'b' is the "slope" of line. It represents change in Y variable for a unit change in X variable.

'a' and 'b' in the equation are called numerical constants because for any given straight line, their value does not change.

If the values of the constants 'a' and 'b' are obtained, the line is completely determined. But the question is how to obtain these values. The answer is provided by the method of Least Squares which states that the line should be drawn through the plotted points in such a manner that the sum of the squares of the deviations of the actual Y values from the computed Y values is the least, or in other words, in order to obtain a line which fits the points best $\Sigma(Y - Y_C)^2$, should be minimum. Such a line is known as the line of 'best fit'.

A straight line fitted by least squares has the following characteristics :

(i) It gives the best fit to the data in the sense that it makes the sum of the squared deviations from the line, $\Sigma(Y - Y_C)^2$, smaller than they would be from any other straight line. This property accounts for the name 'Least Squares'.

(ii) The deviations above the line equal those below the line, on the average. This means that the total of the positive and negative deviations is zero, or $\Sigma(Y - Y_C) = 0$.

(iii) The straight line goes through the overall mean of the data (X, Y).

(iv) When the data represent a sample from a large population the least squares line is a 'best' estimate of the population regression line.

With a little algebra and differential calculus it can be shown that the following two equations, if solved simultaneously, will yield values of the parameters *a* and *b* such that the least squares requirement is fulfilled :

$$\Sigma Y = Na + b\Sigma X$$

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

These equations are usually called the *normal equations.* In the equations ΣX, ΣXY, ΣX^2 indicate totals which are computed from the observed pairs of values of two variables X and Y to which the least squares estimating line is to be fitted and *N* is the number of observed pairs of values.

This will be shown in figure below.

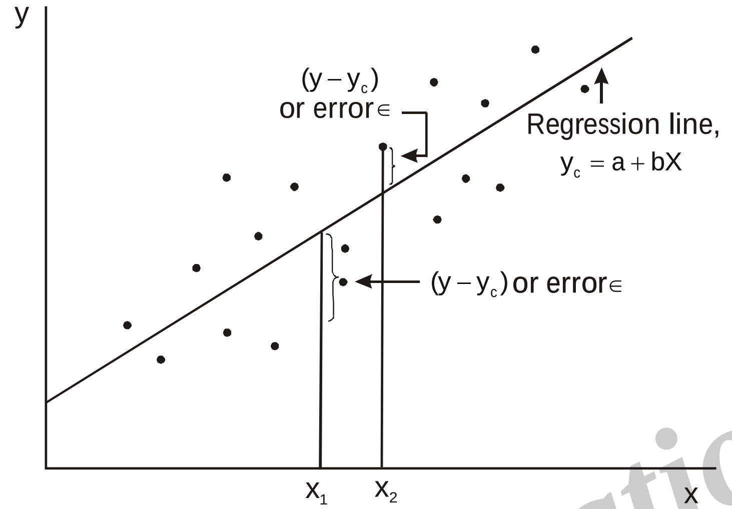


Figure : Regression of *Y* on X: Least Squares

If the parameter estimates be *a* for α and *b* for β, then the line would be

$$Y_c = a + bX$$

Since we seek to minimize $\Sigma(Y-Y_c)^2$, which works out to be $\Sigma(Y-a-bX)^2$, we can find the values of *a* and *b* by applying calculus. This results in a pair of what are called normal equations. The normal equations are :

$$\Sigma Y = na + b\Sigma X$$

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

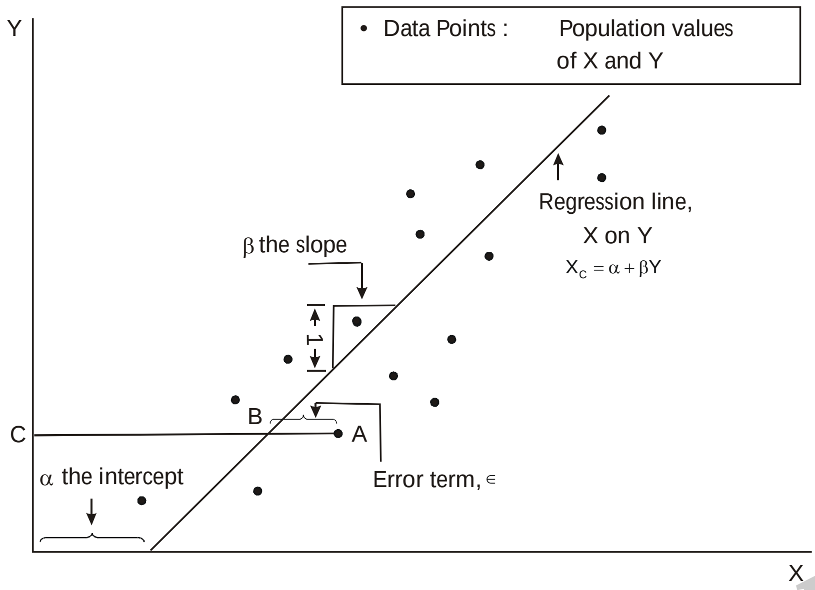
(II) Regression of X on Y

In general, usually X-variable is taken to be independent and the Y-variable as dependent one. However, if the X-variable is treated as the dependent variable and Y as an independent variable, we can also have regression of X on Y. In the regression of X on Y, the population regression model is :

$$X = \alpha + \beta Y + \varepsilon \text{ or } X_C = a + bY$$

in which X is the dependent variable (the variable to be predicted); Y is the independent variable (the predictor variable); α is the population X-intercept: β is the population slope (measured as change in the X variable corresponding to a unit change in Y); and ε is the error term. Here the Y-variable is fixed and the randomness in the X variable comes from the error term, ε.

The population regression model is shown in Figure below.



For a given pair of *X* and *Y* values, represented by a point, say A, the actual value of *X* equal to AC is composed of the non-random part BC (given by the regression line) and the random component AB.

Assumptions

(i) There exists a linear relationship between X and Y variables.

(ii) The values of independent variable Y are fixed while those of dependent variable X are random - with randomness arising from the error term.

(iii) The errors, e, are normally distributed with mean equal to zero, and constant variance a . Further, they are independent in different observations.

Observe here that if X and Y are plotted on the graph on X-axis and Y-axis respectively, then we consider horizontal deviations in the case of regression of X on Y and vertical deviations in the case of regression of Y on X. The estimates of the parameters are given by *a* and *b*, which are obtained from a pair of normal equations given below :

$$\Sigma X = Na + b\Sigma Y$$

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

We calculate the input values and substitute them into the above equations, solve them simultaneously to get *a* and *b*. This yields the regression equation X = a + bY. This equation is then used to get the expected values of X for given values of Y.

Some important points may be noted as follows:

(i) For a given set of paired data, there are two regression lines - one showing regression of Y on X and the second one showing regression of X on Y. One of these is obtained by minimizing the squared vertical deviations and the other by horizontal deviations. As such, they are different lines with separate parameter values.

(ii) The slope parameters of the regression lines are of particular significance. To distinguish, they are designated as b_{yx} and b_{xy}, called regression co-efficient of Y on X and, regression co-efficient of X on Y, respectively.

(iii) For a given set of data, both the regression lines would be either positively sloped or negatively. Thus, both the regression co-efficients would be positive or both would be negative. Positive co-efficients indicate positive correlation and negative co-efficients mean negative correlation. The sign of the other parameter, a, is not important and for the two equations, it may bear same or opposite signs.

(iv) Each of the regression lines passes through the mean values of the variables. When both the regression lines for a given set of paired data are plotted on a graph, their point of intersection yields the mean values of the variables X and Y. Thus, when two regression equations are solved simultaneously, the X and Y values obtained are $\bar{X}$ and $\bar{Y}$, respectively.

(v) The closer are the regression lines to each other, the higher is the degree of correlation between the variables and more away are they from each other, the weaker is the correlation. When the variables are perfectly correlated, the two regression lines coincide. Thus, while usually there are two regression lines for a set of data, when the correlation is perfectly positive or perfectly negative, there would be only one regression line.

The two regression lines in respect of a given set of data are both sloped positively in the case of positive correlation and negatively in the case of negative correlation between the variables. They intersect at $\bar{X}$ and $\bar{Y}$, and their closeness to each other is indicative of the degree of correlation between the variables.

Q10. What are the properties of regression coefficient?

Ans :

(i) $r^2 = b_{xy} * b_{yx}$ In other words, r is the Geometric mean between the two regression coefficients b_{xy} and b_{yx}

(ii) Both the regression coefficients will have the same sign, i.e. either they will be positive or negative. Also, the coefficient of correlation will have the same sign as that of regression coefficients

(iii) The arithmetic mean of the two regression coefficients is greater than the correlation coefficient. In other words, $(b_{xy} + b_{yx})/2 > r$.

(iv) If one regression coefficient is greater than 1, the other has to be less than one. This is an extension of the first property as the product of the two coefficients is equal to square of the correlation coefficient r. Since r lies between -1 and +1, r^2 cannot be greater than 1. Thus, $b_{xy} * b_{yz}$ cannot be greater than 1. Thus, if one regression coefficient is greater than 1, the other has to be less than one .

1.4 Using Regression Lines for Prediction

Q11. Explain briefly about the regression lines used for prediction.

Ans :

In order to predict the value of variable 'X' (in dependent variable) the equation of regression line is preferable.

Following are some of the important points which need to be considered while using regression lines for prediction,

1. The significance of observed sample correlation coefficient r = r(X, Y) is need to be tested. The lines of regression for estimation and prediction can be used when the value of 'r' is significant.

2. Linear model is not a good fit when the 'r' value is not significant. Therefore, lines of regression cannot be used in this case.

3. If 'r' is significant and the linear regression is a good fit for the given data, then it is preferable to use line of regression for estimating 'Y' for given 'X'.

4. Linear regression model should not be used for predicting 'Y' corresponding to far distant value of X because lot of changes may occur in the pattern of relationship between these two variables. Therefore, the predicted value of Y for distant value of X may not be worthy.

Q12. What are the disadvantages of regression analysis?

(OR)

What is the Limitations of Regression Analysis.

Ans : **(Jan.-21, Imp.)**

Despite all utilities, the regression analysis, too, has. various limitations. The following are some of the limitations of regression analysis :

1. **Assumption of Linear Relationship**

   Regression analysis is based on the assumption that there always exists linear relationship between related variables. But the linear type of relationship does not always exist in the field of social sciences. In these fields non-linear or curvilinear relationships are most commonly found.

2. **Assumption of Static Condition**

   While calculating the regression equations a static condition of relationship between the variables is presumed. It is supposed that the relationship has not changed since the regression equation was computed. Such type of assumption has made the regression analysis a static one and hence reduces its applicability in social fields.

3. **Study of Relationship in Prescribed Limits**

   The linear relationship between the variables can only be ascertained within limits. When prescribed limits are crossed, the results become incorrect or inconsistent. Such a relation exists between price and profits. When prices are higher the profits are high to a certain limit. When the prices are abnormally high the sales may go down or some other firms with lower prices may come up with increased supply. Profits may decline due to entry of new firms increasing thereby the supply of the commodity.

PROBLEMS

1. Given the two regression coefficients b_{yx} = 0.4 and b_{yx} = 0.9, find the value of correlation Coefficient.

Sol :

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

$b_{xy} = 0.9,$

$b_{yx} = 0.4 = \sqrt{0.9 \times 0.4} = \sqrt{0.36} = 0.6$

2. If x = 0.85 y and y = 0.89x. Find the coefficient of correlation.

Sol : **(Jan.- 21)**

$x = 0.85y$ (or) $b_{xy} = 0.85$

$y = 0.89x$ (or) $b_{yx} = 0.89$

Coefficient of Correlation (r) $\sqrt{b_{xy} \times b_{yx}}$

$= \sqrt{(0.85)(0.89)} = \sqrt{0.7565}$

$= 0.869$

3. If r = 0.8; σ_x = 2.5, σ_y = 3.5, find bxy and byx.

Sol: **(June-19)**

Given $r = 0.8, \sigma_x = 2.5, \sigma_y = 3.5$

$$b_{xy} = r.\left[\frac{\sigma_x}{\sigma_y}\right] = [0.8]\left(\frac{2.5}{3.5}\right)$$

$\therefore \quad b_{xy} = 0.8 \times 0.7142 = 0.57136$

$$b_{yx} = r.\left[\frac{\sigma_y}{\sigma_x}\right] = [0.8]\left[\frac{3.5}{2.5}\right]$$

$\therefore \quad b_{yx} = (0.8)\ (1.4) = 1.12.$

4. Given that the means of two variables X and Y are 68 and 150, their standard deviations are 2.5 and 20 respectively and the coefficient of correlation between them is +0.6, write down the equation of X and Y.

Sol :

Regression Equation of X on Y.

$= X - \bar{X} = r = \frac{\sigma x}{\sigma y}\left(Y - \bar{Y}\right)$

$= \bar{X} = 68, \bar{Y} = 150,$

$\sigma x = 2.5, \sigma y = 20, r = 0.6$

$= X - 68 = 0.6 \frac{2.5}{20}$

$= (Y - 150) \Rightarrow X - 68$

$= 0.075 Y - 11.25$

$\Rightarrow X = 0.075 Y - 11.25 + 68$

$\Rightarrow X = 0.075 Y + 56.75$

5. For some bivariate data, the following results were obtained. The mean value of X = 53.2 The mean value of Y = 27.9, the regression coefficient of Y on X = –01.5, and the regression coefficient of X on Y = – 0.2 find the most probable value of Y when X = 60.

Sol :

Regression Equation of Y on X

$Y - \bar{Y} = r\frac{\sigma y}{\sigma x}\left(X - \bar{X}\right) = \bar{x} = 53.2,$

$\bar{Y} = 27.9 r\frac{\sigma y}{\sigma x} = 1.5$

$= Y - 27.9 = -1.5 (X - 53.2)$

$\Rightarrow Y - 27.9 = -1.5 (X - 53.2)$

$\Rightarrow Y - 27.9 = -1.5 X + 79.8 + 27.9$

$\Rightarrow Y = 107.7 - 1.5 X$

When X = 60, Y will be 107.7 – 1.5(60)

$= 107.7 - 90 = 17.7$

6. If $\gamma = 0.6$, $\sigma x = 1.5$ and $\sigma y = 2$, Find the bxy and byx.

Sol: **(June-18)**

Regression equation x on y co-efficient is bxy

$$bxy = \gamma \frac{\sigma_x}{\sigma_y}$$

there $\gamma=0.6$, $\sigma x=1.5$ and $\sigma y=2$

γ = Coarctation co-efficient

σ_x = Standard Deviation of 'x' series

σ_y = Standard Deviation of 'y' series

$$b_{xy} = 0.6 * \frac{1.5}{2}$$

0.6 * 0.75

$\boxed{b_{xy} = 0.45}$

- Regression equation y on x co-efficient is byx

$$byx = \gamma \frac{\sigma y}{\sigma x} = 0.6 * \frac{2}{1.5}$$

$0.6*1.333 = \boxed{byx = 0.8}$

7. From the following data obtain the two regression equations and calculate the correlation co-efficient.

x	2	4	6	8	10	12	14	16	18
y	18	16	20	24	22	26	28	32	30

Calculate the value of y when x = 6.2

Sol : **(Jan.- 21, June-19)**

(i) X on Y

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

(ii) Y on X

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$\bar{X} = \frac{\Sigma x}{n} = \frac{90}{9} = 10$$

$$\bar{Y} = \frac{\Sigma y}{n} = \frac{216}{9} = 24$$

Calculation of Regression Equation and Correlation Coefficient

X	Y	$x - \bar{x}$ x	$y - \bar{y}$ y	x^2	y^2	xy
2	18	–8	–6	64	36	48
4	16	–6	–8	36	64	48
6	20	–4	–4	16	16	16
8	24	–2	0	4	0	0
10	22	0	–2	0	4	0
12	26	2	2	4	4	4
14	28	4	4	16	16	16
16	32	6	8	36	64	48
18	30	8	6	64	36	48
90	216	0	0	240	240	228

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{228}{240} = 0.95$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{228}{240} = 0.95$$

Equation X on Y

$X - \bar{X} = b_{xy}(Y - \bar{Y})$

$X - 10 = 0.95\,(X - 24)$

$X = 0.95\,Y - 22.8 + 10$

$X = 0.95\,Y - 12.8$

Equation Y on X

$Y - \bar{Y} = b_{yx}(X - \bar{X})$

$Y - 24 = 0.95\,(X - 10)$

$Y - 24 = 0.95,\ X - 95 + 24$

$Y = 0.95\,x - 14.5$

Regression equation y on x = 6.2

$y = 0.95x - 14.5$

$y = 095(6.2) - 14.5$

$y = -8.61$

8. Given :

$\Sigma x=56$, $\Sigma y=40$, $\Sigma x^2=524$, $\Sigma y^2=256$, $\Sigma xy=364$, $N = 8$

(i) Find the two Regression equations and

(ii) The Correlation Coefficient.

Sol: **(June-18, Imp.)**

We have

$$\bar{X} = \frac{\Sigma x}{N} = \frac{56}{8} = 7; \quad \bar{y} = \frac{\Sigma y}{N} = \frac{40}{8} = 5$$

byx = co-efficient of regression of y on x = $\frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N(\Sigma y^2) - (\Sigma x)^2}$

$$\frac{8(364)-(56)(40)}{8(524)-(56)^2} = \frac{2912-2240}{4192-3136} = \frac{672}{1056}$$

$$\boxed{byx = 0.6363}$$

bxy = co-efficient of regression of x on y = $\frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N(\Sigma y^2) - (\Sigma y)^2}$

$$\frac{8(364)-(56)(40)}{8(256)-(40)^2} = \frac{2912-2240}{2048-1600} = \frac{674}{448} = \boxed{bxy = 1.504}$$

(i) **Two Regression equations**

Regression equation x on y

$(x-\bar{x}) = b_{xy}(y-7)$

(x –7) = 1.504 (y – 5)

(x – 7) = 1.504 (y) – 1.504 (5)

x = 1.504 (y) – 7.522 + 7

x = 1.504 (y) – 0.522 .... (1)

Regression equation y on x

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

(y – 5) = 0.6363 (x – 7)

y – 5 = 0.6363 (x) – 0.6363(7)

y = 0. 6363 (x) – 4.4541 + 5

y = 0.6363 (x) + 0.5459 ....(2)

(ii) **The correlation co-efficient γ_{xy} between x and y is given by**

$$\gamma_{xy}^2 = b_{yx}.b_{xy} = (0.6363)(1.504)$$

$$r_{xy}^2 = 0.9569$$

$$\boxed{\gamma_{xy} = 0.9782}$$

---

9. Following are the marks in Statistics and English in an Annual Examination.

Particular	Statistics (X)	English (Y)
Mean	40	50
Standard Derivation	10	16
Co-efficient Correlation	0.5	

(i) Estimate the score of English, when the score in Statistics is 50.

(ii) Estimate the score of statistics, when the score in English is 30.

Sol: **(June-19, Imp.)**

Given mean of X denoted as $\bar{X}$ = 40.

Given mean of Y denoted as $\bar{Y}$ = 50.

SD of X denoted as σ_x = 10.

SD of Y denoted as σ_y = 16.

Coefficient of correlation denoted as r = 0.5

Regression Equation X on Y

$$[X - \bar{X}] = [r]\left[\frac{\sigma x}{\sigma y}\right][y - \bar{y}]$$

$$X - 40 = [0.5]\left[\frac{10}{16}\right][Y - 50]$$

$$X - 40 = [0.5]\ [0.625]\ [Y - 50]$$

$$X - 40 = [0.3125]\ [Y - 50]$$

$$X - 40 = 0.3125y - 15.625$$

$$X = 0.3125y - 15.625 + 40$$

$$X = 0.3125y + 24.375$$

Regression Equation Y on X

$$[Y - \bar{Y}] = [r]\left[\frac{\sigma y}{\sigma x}\right][X - \bar{X}]$$

$$[Y - 50] = [0.5]\left[\frac{16}{10}\right][X - 40]$$

$$Y - 50 = [0.5]\ [1.6]\ [X - 40]$$

$$Y - 50 = (0.8)\ (X - 40)$$

$$Y - 50 = 0.8X - 32$$

$$Y = 0.8\ X - 32 + 50$$

$$Y = 0.8\ X + 18$$

Estimation of English (Y) when Statistics (X) is 50

$$Y = 0.78\ X + 18$$

$$= 0.8\ (50) + 18$$

$$= 40 + 18$$

$$\therefore Y = 18 \text{ marks.}$$

Estimation of statistics (X) when English (Y) is 30

$$X = 0.3125\ Y + 24.375$$

$$= 0.3125(30) + 24.375$$

$$= 9.375 + 24.375$$

$$X = 33.75 \text{ marks.}$$

10. You are given the following information about advertisement expenditure and sales :

Particulars	Adv. Exp (X) (` Crores)	Sales (Y) (` Crores)
Mean	**20**	**120**
S.D	**5**	**25**
Correlation coefficient	**0.8**	

(i) Calculate two regression equations

(ii) Find likely sales when Adv. Expenses is ` 25 Crores

(iii) What should be the Adv. Budget if the company wants to attain sales target of ` 150 crores.

Sol :

Let, the variable x represent advertisement expenses and y represent sales (in ` crores). Then, we have,

$$\bar{x} = 20,\ \bar{y} = 120,\ \sigma_x = 5,\ \sigma_y = 25,\ r = 0.8$$

(i) Two Regression Equations

(a) Regression Equation of X on Y

$$x - \bar{x} = r\left[\frac{\sigma_x}{\sigma_y}\right](y - \bar{y})$$

$\Rightarrow$ $x - 20 = 0.8 \left[\frac{5}{25}\right](y - 120)$

$\Rightarrow$ $x - 20 = 0.16 (y - 120)$

$\Rightarrow$ $x - 20 = 0.16 y - 19.2$

$\Rightarrow$ $x = 0.16 y - 19.2 + 20$

$\Rightarrow$ $x = 0.16 y + 0.8$

(b) Regression Equation of Y on X

$$y - \bar{y} = r\left[\frac{\sigma_y}{\sigma_x}\right](x - \bar{x})$$

$\Rightarrow$ $y - 120 = 0.8 \left[\frac{25}{5}\right](x - 20)$

$\Rightarrow$ $y - 120 = 4(x - 20)$

$y - 120 = 4x - 80$

$\Rightarrow$ $y = 4x - 80 + 120$

$\Rightarrow$ $y = 4x + 40$

$\therefore$ The two equations are,

$x = 0.16 y + 0.8$

$y = 4x + 40$

(ii) For proposed advertisement expenditure of ` 25 crores, sales,

$y = 4x + 40$

$= 4(25) + 40$

= ` 140 crore

(iii) To achieve the sales target of ` 150 crores the company should have the advertising budget,

$x = 0.16y + 0.8$

$= 0.16 (150) + 0.8$

= ` 24.8 crore.

11. 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**

Sol :

Regression Equation of X on Y

$$X - \bar{X} = bxy\,(Y - \bar{Y})$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{40}{5} = 8$$

$$\therefore \quad \bar{X} = 6, \quad \bar{Y} = 8$$

Calculation of Regression equation and correlation coefficient

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

Number of Pairs N = 5

Regression Co-efficient

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-26}{20} = -1.3$$

$$\therefore \quad X - \bar{X} = b_{xy}\,(Y - \bar{Y})$$

$$X - 6 = 1.3\,(Y - 8)$$

$$X - 6 = 1.3\,Y + 10.4$$

$$X = -1.3Y + 10.4 + 6$$

$$X = 1.3Y + 16.4$$

(or)

$$X = 16.4 - 1.3\,Y$$

Regression Equation of Y on X

$$(Y - \bar{Y}) = b_{yx}\,(X - \bar{X})$$

Regression Coefficient

$$b_{yx} = \frac{\Sigma x^2}{\Sigma y^2} = \frac{-26}{40} = -0.65$$

$\therefore \quad Y - \overline{Y} = b_{yx}(X - \overline{X})$

$Y - 8 = -0.65\,(X - 6)$

$Y - 8 = -0.65x + 3.9$

$Y = -0.65x + 3.9 + 8$

$Y = -0.65x + 11.9$

or

$Y = 11.9 - 0.65x$

---

12. The correlation co-efficient between x and y is (r) = 0.6, $\sigma_x = 1.5$ $\sigma_y = 2$, $\overline{X} = 10$ and $\overline{Y} = 20$, find the two Regression Equations.

Sol :

Given that,

$\bar{x} = 10,\ \bar{y} = 20,\ \sigma_x = 1.5,\ \sigma_y = 2,\ r = 0.6$

(i) Regression Equation of X on Y

$$x - \bar{x} = r\left[\frac{\sigma_x}{\sigma_y}\right](y - \bar{y})$$

$\Rightarrow \quad x - 10 = 0.6\left[\frac{1.5}{2}\right](y - 20)$

$\Rightarrow \quad x - 10 = 0.45\,(y - 20)$

$\Rightarrow \quad x - 10 = 0.45y - 9$

$\Rightarrow \quad x = 0.45\,y - 9 + 10$

$\Rightarrow \quad x = 0.45\,y + 1$

(ii) Regression Equation of Y on X

$$y - \bar{y} = r\left[\frac{\sigma_y}{\sigma_x}\right](x - \bar{x})$$

$\Rightarrow \quad y - 20 = 0.6\left[\frac{2}{1.5}\right](x - 10)$

$\Rightarrow \quad y - 20 = 0.799\,(x - 10)$

$\Rightarrow \quad y - 20 = 0.799\,x - 7.99$

$\Rightarrow \quad y = 0.799\,x - 7.99 + 20$

$\Rightarrow \quad y = 0.799x + 12.01$

$\therefore$ The two regression equations are,

$x = 0.45\,y + 1$

$y = 0.799x + 12.01$

Short Question and Answers

1. Define Regression

Ans :

Regression analysis which confines itself to a study of only two variables is called simple regression. The regression analysis which studies more than two variables at a time is called multiple regression. In the simple regression analysis there are two variables-one of which is known as 'independent variable' or 'regressor' or 'predictor'. On the basis of the values of this variable the values of the other variable are predicted. The other variable whose values are predicted is called the 'dependent' or 'regressed' variable.

Definitions

1. "Regression is the measure of the average relationship between two or more variables in terms of the original units of the data."
2. **According to Morris Hamburg** The term 'regression analysis' refers to the methods by which estimates are made of the values of a variable from a knowledge of the values of one or more other variables and to the measurement of the errors involved in this estimation process."
3. **According to Taro Yamane** "One of the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related causally, is regression analysis."
4. **According to YaLum Chou** "Regression analysis attempts to establish the 'nature of the relationship between variables that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting."

2. Importance of regression analysis

Ans :

1. Regression analysis helps in establishing a functional relationship between two or more variables. Once this is established it can be used for various advanced analytical purposes.
2. Since most of the problems of economic analysis are based on cause and effect relationship, the regression analysis is a highly valuable tool in economics and business research.
3. This can be used for prediction or estimation of future production, prices, sales, investments, income, profits and population which are indispensable for efficient planning of an economy and are of paramount importance to a businessman or an economist.
4. Regression analysis is widely used in statistical estimation of demand curves, supply curves, production functions, cost functions, consumption functions, etc. Economists have discovered many types of production functions by fitting regression lines to input and output data.

3. Linear Regression.

Ans :

Linear regression is a form of regression which is used for modeling the relationship between scalar variables like X and F under linear regression, linear functions are used to model the data and the unknown parameters, of models are estimated from the data. Hence, these models are known as linear models.

4. Non-Linear Regression

Ans :

In the non-linear regression the explained variable (dependent variable) changes at varying rate with a given change in the explaining variable (independent variable). It is also known as Curvilinear regression. Under non linear regression, the observational data are modeled by a function i.e., a non linear blend of model parameters and depends on one or more independent variable.

Method of successive approximations are used for fitting the data. The data in non linear regression contains of error free independent variable 'X' and its relatively observed dependent variable 'Y.

Example

The output of rice increases rapidly with the application of the initial dose of fertilizer; there after it increases at a falling rate. The relationship in such case, when shown on graph will yield a 'curve'.

5. Compare and contrast linear and non-linear regression.

Ans :

The differences between linear and non-linear regression are as follows,

S.No.	Basis	Linear Regression	Non-Linear Regression
1.	**Meaning**	Linear regression is a form of regression which is used for modelling the relation ship between a scalar variable 'X' and 'Y'.	Non linear regression is a type of regression, where the observational data are modeled by a function i.e., a non linear blend of model parameters.
2.	**Curve**	If the regression curve is a straight line, then the regression is termed as linear regression.	If the curve of the regression is not a straight line. then the regression is termed as curved or non-linear regression.
3.	**Model form**	Under this, the parameters are considered as linear combinations.	Under this, the parameter are considered as functions
4.	**Solution**	Under linear regression, the solution for parameters is represented as closed form	Under non linear regression it is necessary for parameters to be solved repeatedly by using optimization algorithms.
5.	**Uniqueness**	The solution under linear regression is unique.	The Sum of the Squared Errors (SSE) may not be appear as unique.
6.	**Parameters estimation**	In case of uncorrelated errors, estimation Parameters are unbiased.	Incase of uncorrelated errors, estimation of Parameters are usually biased.

6. What do you mean by line of regression?

Ans :

In a bi-variate distribution, if the variables are related then the points when plotted in the scatter diagram will lie near a straight line which is called the line of regression and the regression is said to be linear regression. If points lie on some non-linear curve then the regression is said to be curvilinear regression.

7. What are the properties of regression coefficient?

Ans :

(i) $r^2 = b_{xy} * b_{yx}$ In other words, r is the Geometric mean between the two regression coefficients b_{xy} and b_{yx}

(ii) Both the regression coefficients will have the same sign, i.e. either they will be positive or negative. Also, the coefficient of correlation will have the same sign as that of regression coefficients

(iii) The arithmetic mean of the two regression coefficients is greater than the correlation coefficient. In other words, $(b_{xy} + b_{yx})/2 > r$.

(iv) If one regression coefficient is greater than 1, the other has to be less than one. This is an extension of the first property as the product of the two coefficients is equal to square of the correlation coefficient r. Since r lies between -1 and +1, r^2 cannot be greater than 1. Thus, $b_{xy} * b_{yz}$ cannot be greater than 1. Thus, if one regression coefficient is greater than 1, the other has to be less than one .

8. What is the Limitations of Regression Analysis?

Ans :

Despite all utilities, the regression analysis, too, has. various limitations. The following are some of the limitations of regression analysis :

1. Assumption of Linear Relationship

Regression analysis is based on the assumption that there always exists linear relationship between related variables. But the linear type of relationship does not always exist in the field of social sciences. In these fields non-linear or curvilinear relationships are most commonly found.

2. Assumption of Static Condition

While calculating the regression equations a static condition of relationship between the variables is presumed. It is supposed that the relationship has not changed since the regression equation was computed. Such type of assumption has made the regression analysis a static one and hence reduces its applicability in social fields.

3. Study of Relationship in Prescribed Limits

The linear relationship between the variables can only be ascertained within limits. When prescribed limits are crossed, the results become incorrect or inconsistent. Such a relation exists between price and profits. When prices are higher the profits are high to a certain limit. When the prices are abnormally high the sales may go down or some other firms with lower prices may come up with increased supply. Profits may decline due to entry of new firms increasing thereby the supply of the commodity.

9. Define the principle of least squares and standard error of estimate.

Ans :

Principle of Least Squares

The principle of least squares consists of minimizing the sum of the squares of the residuals or the errors of estimates, i.e., the deviations between the given observed values of the variable and their corresponding estimated values as given by the line of best fit.

Standard Error of Estimate

The standard error of estimate is a measure of the accuracy of predictions. The estimates obtained by using the regression equations may not be prefect.

10. What are the various types of regression variable?

Ans :

The various types of regression variables are as follows,

(i) Independent Variable

The variable which influences the values of the other variable or which is used for predicting the value of the other variable is called independent variable.

(ii) Dependent Variable

The variable whose value is influenced or is to be predicted is called dependent variable. Regression test generates lines of regression of the two variables which helps in estimating the values. Lines of regression of y on x is the line which gives the best estimate for the value of y for any specified value of x. Similarly, line of regression of x on y is the line which gives the best estimate for the value of x for any specified value of y.

Exercise Problems

1. Given the two regression Coefficient X on Y = +0.542 and Y on X = +0.905. Calculate the coefficient of correlation between X and Y.

 [Ans : +0.70]

2. If the regression Coefficient of X on Y = 0.847 and Y on X = +0.732, find coefficient of correlation.

 [Ans : +0.787]

3. If Regression equation of X on Y = 0.268 and of Y on X = 0.5, find coefficient of correlation.

 [Ans : +0.259]

4. Fit a straight line regression equation of Yon X from the following data.

X	10	12	13	16	17	20	25	29
Y	10	12	24	27	29	33	37	42

 [Ans : Y = 1.6 X- 1.65]

5. Find the two regression equations from the following data :

X	1	2	3	4	5
Y	2	3	5	4	6

 [Ans : X = 0.9 Y -0.6; Y = 0.9 X + 1.3]

6. From the following data, obtain the two regression lines.

X	2	6	8	11	13	13	13	14
Y	8	6	10	12	12	14	14	20

 [Ans : X = 0.8125Y + 0.25; Y = 0.8125 X + 3.875]

7. From the following data, find out two regression equations.

X	32	46	57	65	55	72	80	67	75
Y	100	95	87	110	95	92	88	90	85

 [Ans : X = –0.8134 Y + 137.0981; Y =-0.21X +106.40]

Choose the Correct Answers

1. The regression equation of X on Y gives [a]
   (a) The most probable values of X for given values of Y
   (b) The most probable values of Y for given values of X
   (c) Either of the two
   (d) None

2. The concept of regression was given by [b]
   (a) Sir Francis Galton in 1807 (b) Sir Francis Galton in 1877
   (c) Sir Hanrey Fayol in 1854 (d) Charles Babbage in 1867.

3. Regression is : [c]
   (a) Measures of average relationship between two more variables
   (b) To find a relation between two or more variables that are related casually.
   (c) Both of the above
   (d) None of these.

4. Number of observations in regression analysis is considered as [d]
   (a) Degree of possibility (b) Degree of average
   (c) Degree of variance (e) Degree of freedom

5. If all conditions or assumptions of regression analysis simple regression can give [c]
   (a) Dependent estimation (b) Independent estimation
   (c) Reliable estimates (d) Unreliable estimates

6. In Regression Analysis, testing of assumptions if these are true or not is classified as [d]
   (a) Weighted analysis (b) Average analysis
   (c) Significance analysis (d) Specification analysis

7. A process by which we estimate the value of dependent variable on the basis of one or more independent variables is called: [b]
   (a) Correlation (b) Regression
   (c) Residual (d) Slope

8. The method of least squares dictates that we choose a regression line where the sum of the square of deviations of the points from the lie is: [b]
   (a) Maximum (b) Minimum
   (c) Zero (d) Positive

9. If one regression coefficient is greater than one, then other will he: [c]
   (a) More than one (b) Equal to one
   (c) Less than one (d) Equal to minus one

10. The dependent variable is also called: [d]
    (a) Regressand variable (b) Predictand variable
    (c) Explained variable (d) All of these

Fill in the Blanks

1. The sign of regression coefficient is ....as that of correlation coefficient.
2. The regression analysis measures ....between X and Y.
3. The purpose of regression analysis is to study .... between variation.
4. When one regression coefficient is positive the other would be ....
5. Lines of regression are ...... if r = 0, and they are ...... if r = ± 1
6. The farther the two regression lines cut each other the ...... be the degree of correlation.
7. If the regression coefficient of X on Y and Y on X are –0.4 and –0.9 respectively then the correlation coefficient is ....
8. If one of the regression coefficient is ....unity the other must be .... unity.
9. The statistical tool with the help of which we estimate the .... of one variable from the .... value of another variable is called ....
10. Both the regression coefficients cannot ....1.
11. If both the regression coefficients are negative , the correlation coefficient would be ....
12. The variable we predict is called the ....
13. The regression analysis help us to study the ....relationship between the variables.
14. The square root of .... coefficients gives us the value of correlation coefficient.

ANSWERS

1. Same
2. Average relationship
3. Dependence
4. Positive
5. Perpendicular, same
6. Lesser
7. –0.6
8. > , <
9. Unknown variable, known, regression
10. Exceed
11. Negative
12. Dependent variable
13. Nature of
14. Regression

UNIT II	**INDEX NUMBERS :** Introduction - Uses - Types - Problems in the Construction of Index Numbers - Methods of Constructing Index Numbers - Simple and Weighted Index Number (Laspeyre - Paasche, Marshall – Edgeworth) - Tests of Consistency of Index Number: Unit Test - Time Reversal Test - Factor Reversal Test - Circular Test - Base Shifting - Splicing and Deflating of Index Numbers.

2.1 Index Numbers

2.1.1 Introduction

Q1. Define Index Numbers.

Ans : **(Jan.-21, June-19, Imp.)**

Meaning

An Index, simply stated, is an indicator. It indicates the broad change in a given phenomenon. It is relative and indicates changes in the level of the given phenomenon in course of time, across geographical locations or in respect of any other characteristic. For example, the BSE Sensex (stands for Bombay Stock Exchange Sensitive Index) is an index of movement in stock prices. It tells us, very broadly, whether the prices of stocks of various companies have gone up or moved down on the Bombay Stock Exchange. Similarly, the Wholesale Price Index (WPI) gives us an indication of whether the general price level in the economy is going up or falling down. Similarly, Index numbers (or indices) can be constructed to measure any phenomenon, such as Industrial production, Number of road accidents, cost of Living or any other activity. Thus, Index Numbers are barometers measuring change in the level of a phenomenon.

Definitions

(i) **According to Croton and Cowden** Different Experts have defined Index Numbers in different words. Some of the definitions are stated as below.

(ii) **According to Speiges** "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

(iii) **According to Weldon** "An Index Number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographical location or any the characteristic."

(iv) **According to Edgeworth** "An Index Number is a statistical device for indicating the relative movements of data where measurement of actual movement is difficult or incapable of being made."

(v) **According to Bowley** "Index Number shows by its variations the changes in magnitude which is not susceptible either of accurate measurement in itself or of direct valuation in practice."

"Index Numbers are used to measure the change in some quantity which cannot be observed directly, which we know to have a definite influence on many other quantities which we can so observe, tending to increase all or diminish all, while this influence is concealed by the action of many causes affecting the separate quantities in various ways."

On the basis of the above definitions, the following points can be made:

1. Index Numbers are a measurement device
2. They do not measure or state the actual level attained by the phenomenon being studied. They measure the change in the phenomenon being studied.
3. The situations for which Index Numbers are used for comparison are not restricted in any manner. It can be a comparison of two time periods, two geographical locations, two groups of people or any other phenomenon.

4. Index Numbers are the result of a numerical calculation. They do not have any units such as kgs or rupees.

5. Index Numbers are relative terms and hence, they are normally expressed in percentage terms.

2.1.2 Uses

Q2. What are the uses of index numbers ?

(OR)

Explain the importance of index numbers.

Ans : **(June-19, June-18, Imp.)**

1. Economic Barometers

Index Numbers can be constructed for any phenomenon for which quantitative information is available. They capture the various changes taking place in the general economy and business activities. They provide a fair view of the general trade, the economic development and business activity of the country. Thus, they are aptly termed as 'Economic barometers'.

2. Study of Trend

Index Numbers study the relative changes in the level of a phenomenon over different periods of time. They are especially useful for study of general trend of the general trend for a group phenomenon in a time series data.

3. Policy Formulation

Index Numbers are indispensable for any organization in efficient planning and formulation of executive decisions. For example, the Dearness Allowance payable to employees is determined on the basis of Cost of living index numbers. Similarly, Psychiatrists use Intelligence Quotients to assess a child's intelligence in relation to his/her age, which further can be used in framing the education policy.

4. Deflation

Index Numbers can be deflated to find out the real picture pertaining to the phenomenon. For example, we can know if the real income of employees has been increasing by deflating the nominal wages with the help of index numbers.

5. Forecasting

Index Numbers provide valuable information that aids forecasting. For example an Index of sales, along with related indices such as cost of living index, helps in forecasting future demand and future sales of business.

6. Measurement of Purchasing Power

The cost of Living Index helps us in finding out the intrinsic worth of money. It is one of the key indicators touching the life of the common man.

7. Simplicity

Index Numbers eliminate the clutter of large numbers and complicated calculations, providing the underlying information in a manner that is simple and easy to underlying information in a manner that is simple and easy to understand. For example, many people may not understand the dynamics of stock markets, but they can follow the movements of the BSE Sensex with ease.

Q3. What are the characteristics of Index Numbers ?

Ans : **(June-19, Imp.)**

The following are the main characteristics of index numbers :

i) Index Numbers are Specialized Averages.

ii) Index Numbers as Percentages and Measure of Relative Change.

iii) Basis of Comparison.

iv) Universal Applications.

(i) Index Numbers are Specialized Averages: According to L.R. Connor, "In its simplest form an index number represents a special case of an average, generally a weighted average compiled from a sample of items judged to be representative of the whole."

In simple average the data treated have same units of measurement but index number average variables having different units of measurement. We may find an index number for a group of items including food, clothing, rent, cooking gas etc. which are all measured in different units.

(ii) Index Numbers as Percentages and measure of Relative Change: Index numbers are expressed in percentages but the word or symbol for percentage is never used. They measure relative changes in price over years with reference to some period called Base Period or Base Year. If the index number in 2000 is 150 with reference to 100 in 1995. It will mean that there is 50% increase in price in 2000 as compared to 1995. Since the index numbers are quantitative expressions of relative changes, they are expressed in numbers.

(iii) Basis of Comparison : Index numbers are used to make comparisons over different time periods with reference to base period.

(iv) Universal Applications : They are applied universally to ascertain different types of changes in different sectors of the economy. These changes may be with regard to prices, standard of living of the people, industrial or agricultural production, national or per capita income, exports or imports etc.

2.2 Types of Index Numbers

Q4. Explain the various types of index numbers.

Ans : **(June-19, Imp.)**

An Index is an indicator. Index Numbers can be constructed for study of any phenomenon, if such phenomenon is capable of being quantified. In relation to data pertaining to business and economy, index numbers may be classified into the following.

1. Price Index Numbers

Price Index numbers measure the changes in prices. They are the most common indices. There are various methods by which the indices can be calculated. Most of the discussion in this chapter shall relate to price index numbers.

2. Quantity Index Numbers

Quantity Index Numbers study the changes in the volume of goods produced, consumed or transacted. The Index of Industrial Production is an example of a quantity index. However Quantity Index numbers are not very popular.

3. Value Index Numbers

Value Index numbers are rarely used. They study the change in total value, rather than mere change in prices of changes in quantities. They may have to be supplemented with price and quantity indices.

2.3 Problems in the Construction of Index Numbers

Q5. What are the problems involved in construction of index numbers ? Explain.

Ans : **(Imp.)**

The following problems are mainly faced in the construction of index numbers :

i) Definition of the Purpose,

ii) Selection of the Base Period,

iii) Selection of Items,

iv) Selection of Sources of Data and Collection of Data,

v) Selection of Average,

vi) System of Weighting.

(i) Definition of the Purpose

There are no all purpose index numbers. Therefore, before constructing an index number the specific purpose, i.e., objective for which it is designed must be clearly and rigorously defined. Haberler has rightly said "Different index numbers are constructed to fulfil different objectives and before setting to construct a particular number one must clearly define one's objective of study because it is on the objective of the study that the nature and format of the index number depends".

(ii) Selection of the Base Period

Selection of the proper Base Period is an important factor in the construction of index numbers. Base Period is a reference point with which changes in other periods are measured. About the selection of Base Period, following observations may be noted.

Morris Hamburg observes, "It is desirable that the Base Period be not too far away in times from the present. The further away we move from the Base Period, the dimmer are our recollections of Economic conditions prevailing at that time. Consequently comparisons with these remote periods tend to lose significance and to become rather tenuous in meaning."

According to George Simpson and Fritz Katka, "Since practical decisions are made in terms of index numbers, and economic practices so often are a matter of the short run, we wish to make comparisons between a base which lies in the same general economic framework as the years of immediate interest. Therefore, we choose a base relatively close to the years being studied." In the fast changing world of today the base year should not be more than a decade old. There is a psychological reason also for taking a recent period as base.

(iii) Selection of Items

Selection of items or 'Regimen' or 'Basket'. In any index number neither it is possible nor necessary to include all the items or commodities. Each index number tries to measure changes pertaining to a particular group.

Selection of items also depends on the purpose of index number. Moreover, items selected should be such as are widely consumed. The items selected for an index number should be relevant, representative, reliable and comparable. In general the larger is the number of items, the lesser will be the chances of error in the average. Since we cannot include a very large number of items. A compromise is always required between the number of items and the reasonable standard of accuracy. We must, however, have manageable number of items and should also aim at reasonable standard of accuracy.

(iv) Selection of Sources of Data and Collection of Data

For sources of data and collection of data we are mainly concerned with she prices. Use of wholesale prices or retail prices depends on the objective of study. Price quotations should be obtained from important markets. In order to ensure better results, it is advisable to take a standard price which implies representative price of a commodity for whole interval under consideration.

(v) Selection of Average

Since index numbers measure the relative changes, Geometric mean should be the best average but due to certain difficulties in calculations with G.M., for all practical purposes Arithmetic mean is used.

(vi) System of Weighting

According to John 1 Giffin, "In simple terms, weighting is designed to give component series an importance in proper relation to their real significance." In order to allow each commodity to have a reasonable influence on the index it is advisable to use a suitable weighting system.

In case of an unweighted index number of prices, all commodities are given equal importance. But in actual practice different commodities need a different degree of importance.

The weights may be according to :

- The value or quantity produced.
- The value or quantity consumed.
- The value or quality sold.

Selection of an Appropriate Formula

A large number of formulae have been devised for constructing the index. The problem very often is that of selecting the most appropriate formula. The choice of the formula would depend not only on the purpose of the index but also on the data available.

Prof. Irving Fisher has suggested that an appropriate index is that which satisfies time reversal test and factor reversal test. Theoretically. Fisher's method is considered as "ideal" for constructing index number. However, from a practical point of view there are certain limitations of this index which shall be discussed later. As such, no one particular formula can be regarded as the best under all circumstances. On the basis of this knowledge of the characteristics of different formulae, a discriminating investigator will choose technical methods adapted to his data and appropriate to his purposes.

2.4 Methods of Constructing Index Numbers

2.4.1 Simple and Weighted Index Number

Q6. What are the various methods of Constructing Index Numbers?

Ans : **(Imp.)**

The various methods of constructing index numbers are shown in the following figure :

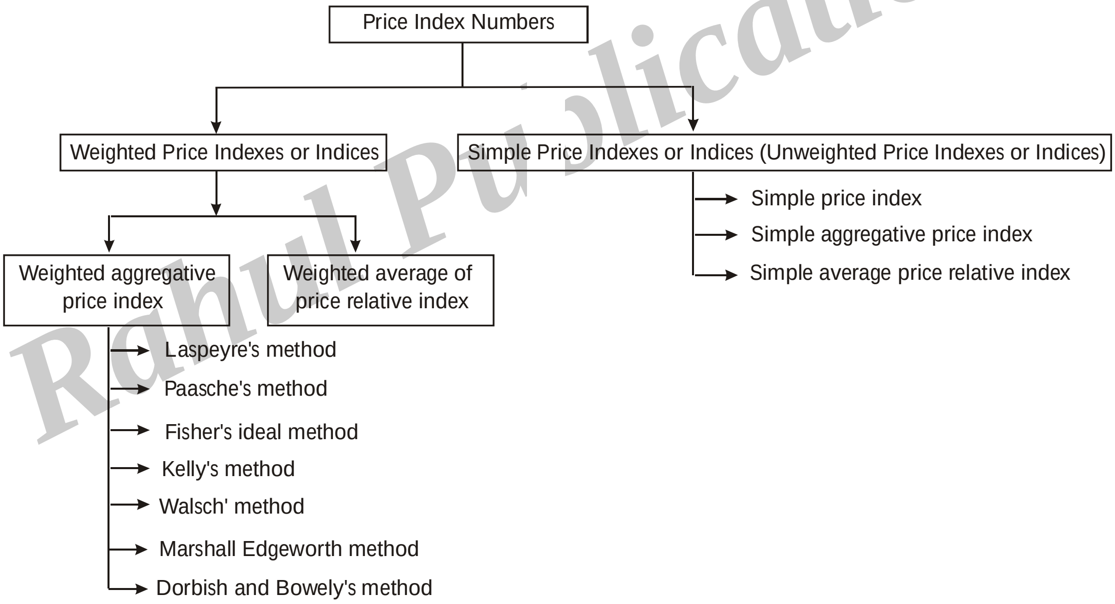


Fig : Types and Methods of Price Index Numbers

1. Weighted Price Indexes

At the time of constructing the weighted price indexes or indices, the rational weights are allocated in an explicit manner. These rational weights show the relative significance of items or commodities which are related with the computation of an index. Quantity weights and value weights are used in this weighted indexes or indices. Weighted price indexes or indices are further divided into two types as follows,

(a) Weighted aggregate price index

(b) Weighted average of price relative index.

(a) Weighted Aggregate Price Index

In a weighted aggregate price index, certain weight is assigned to each and every commodity or item of group in accordance with its significance. This helps in gathering more information and improving accuracy of the estimates.

The following methods are used in weighted aggregate price index:

(i) Laspeyre's method

(ii) Paasche's method

(iii) Fisher's ideal method

(iv) Kelly's method

(v) Walsch's method

(vi) Marshal Edgeworth's method

(vii) Dorbish and Bowley's method.

(b) Weighted Average of Relatives Method

The basis methodology of calculating the price relatives is same as in case of simple average of relatives method. However, instead of calculating simple average weights are assigned to price relatives and a weighted average is calculated. Such weighted average can be arithmetic mean or geometric mean of the weighted price relatives. The following steps have to be followed.

Case (i) If weighted Arithmetic Mean is used.

Step 1:

Calculate Price Relatives P for each item, $p = (P_1/P_0) \times 100$

Step 2:

Calculate weights with which price relatives are to be multiplied.Normally, the value of the item in the base year (i.e. P_0q_0) is taken as weight. How ever, weights can be P_0q, p_1q_0 or p_1q_1. Weights are denoted by V.

Step 3 :

Calculate weighted price relatives by multiplying the price relatives with their corresponding weights. In other words, Calculate PV.

Step 4 :

Add weighted price relatives obtained in step 3. Denote as $\sum PV$

Step 5 :

Calculate the sum of weights. This is denoted by $\sum V$

Step 6 : Weighted Average of Price Relatives $= p_{0.1} = \dfrac{\Sigma PV}{\Sigma V}$

Case (ii) If Geometric mean is to be used.

Step 1 :

Calculate Price relatives P for each item. $P = (P_1/p_0) \times 100$

Step 2 :

Calculate the logarithm value of P. This is denoted by log P

Step 3 :

Calculate weights for each them. This is denoted by V.

Step 4 :

Calculate V log P for each item by multiplying weight obtained in step 3 with logarithm value of price relatives obtained in step 2.

Step 5 :

Add the weighted logarithm values of Price relatives. Denote it as $\sum(\text{V.log P})$.

Step 6 : Weighted Average to price relatives

$$= P_{0.1} = \text{Antilog}\left[\frac{\sum(\text{V.log P})}{\sum V}\right]$$

2. Unweighted Price Index

(a) Simple Aggregative Method

This method involves aggregation of prices in the current period and expressing the total as a percentage of aggregate of prices in the base period. The following steps are followed.

Step 1:

Calculate $\sum P_1, \sum P_1$ is the sum total of prices of all items in the current year.

Step 2:

Calculate $\sum P_0, \sum P_0$ is the sum total of prices of all items in the base year.

Step 3:

Calculate $P_{0.1}$, $P_{0.1}$, is the price index number of the current year with respect to the base year. It is expressed in percentage terms and calculated as under:

$$P_{0.1} = \left(\frac{\sum P_1}{\sum P_0}\right) \times 100$$

(b) Simple Average of Price Relative Method

Under this method, the price of each item in the current year is expressed as a percentage of its price in the base year. The figure so obtained is called Price Relative. The price relatives are then averaged to calculate the index number for that year. Either Arithmetic mean or Geometric mean can be used for calculation of average of price relatives. The following steps are following.

(a) If Arithmetic mean is used for the purpose of averaging then

Step 1:

Calculate Price relative for each item. Price Relative of an item. Is obtained by the formula $\left(\frac{p_1}{p_0}\right) \times$ 100 Where P1 is price of the item in the current year and P0 is the price of the item in the base year.

Step 2 :

Calculate average of price relatives to obtain the Index Number P0.1

$$P_0.1 = \sum\frac{\left(\frac{p_1}{p_0}\times 100\right)}{N}$$

Where N = Number of Items

(b) If Geometric Mean is used for averaging the price relatives, then

Step 1 :

Calculate Price Relative (p) of each item (p = P1/ p0) × 100

Step 2 :

Calculate Logarithm Value of each Price relative (log p)

Step 3 :

Calculate simple average of logarithm values obtained in step $2\left(\frac{\text{Logp}}{N}\right)$

Step 4 :

Calculate Antilog of value obtained in step 3 Thus

Index Number $P_{0.1}$, = Antilog $\left[\frac{\sum \log p}{N}\right]$

Where $p = \left(\frac{p_1}{p_0} \times 100\right)$

2.4.2 Laspeyre's

Q7. What is Laspeyre's index method ?

Ans :

This method takes the quantities of the commodities in the base period as the weight of that commodity for the purpose of calculating the index numbers. The following steps may be following.

Step 1:

Multiply the current year price (represented by p_0) with the quantities of the base year (q_0) for each commodity.

Step 2:

Add the numbers obtained in step 1. The resultant sum is represented as $\sum p_1 q_0$

Step 3 :

Multiply the prices of base year (represented by p_0) with the quantities of the base year for each commodity.

Step 4 :

Add the numbers obtained in step 3. The resultant sum is represented as $\sum p_1 q_0$

Step 5:

The index number as per Laspeyre's method

$$= P_{0.1} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

2.4.3 Paasche's

Q8. What is Paasche's index method ?

Ans :

Paasche's Method is similar to Laspeyre's method. The only difference is in assignment of weights. As per this method quantities consumed of the commodities in the current year is taken as basis. The following steps need to be followed.

Step 1:

Multiply current year's prices (p_1) with current year's quantities (q_1)

Step 2 :

Add the numbers obtained in step (1). The resultant sum is $\sum p_1q_1$

Step 3 :

Multiply base year's Prices (p0) with current year's quantities (q_1)

Step 4 :

Add the numbers obtained in step (2). The resultant sum is $\sum p_0q_1$

Step 5 :

Index Number as per Paasche's method $= P_{0.1} = \frac{\sum p_1q_1}{\sum p_0q_1} \times 100$

Q9. Explain the comparision of Laspeyre's and Paasche's method ?

Ans :

S.No.	Laspeyre's Index Number	S.No.	Paasche's Index Number
1.	Here, quantity of the base year is assumed to be the quantity of the current year.	1.	Here, quantity of the current year is assumed to be the quantity of the base year.
2.	It has an upward bias i.e. the numerator of the index number is increased due to the assignment of higher weights fixed on the basis of the base year's quantities even though there might have been a fall in the quantity consumed during the current year due to rise, or fall in price and change in in tastes, habits and customs etc. in the current year.	2.	It has a downward bias i.e. the numerator of the index number is decreased due to the assignment of lower weights fixed on the basis of the current year's quantities even though the quantities in the current year might have fallen due to rise or fall in price, or change in habits of consumption.
3.	As the quantity of the base year are used as weights, the influence of price changes on quantities demanded do not get reflected in the index number.	3.	As the quantities of the current year are used as weights, the influence of price changes on quantities demanded get reflected in the index number.
4.	It measures changes in a fixed marked basketof goods and services as the same quantities are used in each period.	4.	It continually updates the quantities to the level of current consumption.
5.	Here, weights remain constant.	5.	Here, weights are determined every time an index number is constructed.

2.4.4 Marshall – Edgeworth

Q10. What is Marshall Edgeworth method ?

Ans :

In this method also both the current year as well as base year prices and quantities are considered. The formula for constructing the Index is :

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

on opening the brackets

$$P_{01} = \frac{\Sigma\, p_1q_0 + \Sigma\, p_1q_1}{\Sigma\, p_0q_0 + \Sigma\, p_0q_1} \times 100$$

It is a simple, readily constructed measure, giving a very close approximation to the results obtained by the ideal formula.

2.4.5 Fisher's Ideal Index

Q11. What is Fisher's Ideal Index ?

Ans :

Fisher's Ideal Index

This is the most popular amongst all weighted aggregative index numbers. It is obtained by calculating the Geometric Mean (G.M) of Laspeyre's and Paasche's index numbers. The formula for calculating fisher's ideal index is an under.

$$P_{0.1} = \left[\sqrt{\frac{\sum p_1q_0 \times \sum p_1q_1}{\sum p_0q_0 \times \sum p_0q_1}}\right] \times 100$$

Reasons for Fisher's Index being called an Ideal Index

Reasons

(i) It gives weightage to both current consumption and base year consumption.

(ii) It is free from upward or downward bias.

(iii) It satisfies both time reversal and factor reversal tests (to be discussed later) .

(iv) It is a Geometric mean of Laspeyre's index and Paasche's index

PROBLEMS

1. From the following data construct an Index number for 1994 taking 1999 as base as per simple Aggregative method.

Commodities	Price in 1993 (Rs)	Price in 1994 (Rs)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40

Sol :

Construction of Price Index

Commodities	Price in 1993 in Rs. (P_o)	Price in 1994 in Rs. (P_1)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40
	ΣP_o = 240	ΣP_1 =350

$$P_{01} = \left(\frac{\Sigma p1}{\Sigma p0}\right) \times 100 \frac{350}{240} \times 100 = 145.83$$

2. From the following data construct an Index number for 1994 taking 1999 as base as per simple Aggregative method.

Year	Commodity					
	A	B	C	D	E	F
1970	45	60	20	50	85	120
1975	55	70	30	75	90	130

Sol :

(a) Arithmetic Mean

Commodity	P_0 (Price in 1970)	P_1 (Price in 1975)	Price Relative P	Log P
A	45	55	55/45× 100 = 122.22	2.0871
B	60	70	116.67	2.0668
C	20	30	150.00	2.1761
D	50	75	150.00	2.1761
E	85	90	105.88	2.0245
F	120	130	108.33	2.0346
			ΣP=753.10	**12.565.2 = Σ Logp**

Step 2 : Average of Price Relatives $= \frac{753.10}{6} = 125.52$

(b) Geometric Mean

$$P_{0.1} = \text{Antilog}\left[\frac{\Sigma \text{Logp}}{N}\right] = \text{Antilog}\left(\frac{12.5652}{6}\right) = \text{Antilog}\ (2.0942) = 124.3$$

3. From the following data calculate a price Index based on price Relatives Method using Arithmetic Mean.

Commodity	A	B	C	D	E	F
Price 2015(Rs.)	45	60	20	50	85	120
Price 2016 (Rs.)	55	70	30	75	90	130

Sol : **(June-18)**

Index Number using Arithmetic mean of price relations.

Commodity	Price in 2015 (P_o) (Rs)	Price in 2016 (P_1) (Rs)	Price Relatives $\frac{P_1}{P_o} \times 100$
A	45	55	$\frac{55}{45} \times 100 = 122.22$
B	60	70	116.666
C	20	30	150
D	50	75	150
E	85	90	105.88
F	120	130	108.33
N = 6		Total	$\Sigma \frac{P_1}{P_2} \times 100 = 753.09$

Arithmetic Mean $P_{01} = \frac{\Sigma \frac{P_1}{P_0} \times 100}{N} = \frac{753.09}{6} = 125.515$

$\boxed{A.M = 125.515}$

4. From the following data construct on Index Number for 2017 taking 2016 as base as per simple aggregative Method.

Commodity	P	Q	R	S	T
Price 2015 (`)	40	60	85	25	30
Price 2017 (`)	60	90	125	30	40

Sol : **(Jan.- 21, Imp.)**

	Price	
Commodity	2015 (P_0)	2017 (P_1)
P	40	60
Q	60	90
R	85	125
S	25	35
T	30	40
Total	$\Sigma p_0 = 240$	$\Sigma p_1 = 350$

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

$$= \frac{350}{240} \times 100 = 1.4583 \times 100$$

$$= 145.83$$

5. Calculate Index number by average price relative method by using arithmetic mean.

Commodity	P	Q	R	S	T
Price 2017	2	6	10	5	12
Price 2018	4	8	15	5	8

Sol : **(Jan.- 19)**

Commodity	P_0	P_1	$P = \frac{P_1}{P_0} \times 100$
P	2	4	200.00
Q	6	8	133.33
R	10	15	150.00
S	5	5	100.00
T	12	8	66.66
			649.99

Average of price relatives using A.M.

$$P_{01} = \frac{\Sigma P}{N} = \frac{649.99}{5} = 129.998 = 130$$

6. From the following data compute Laspeyre's Index number for 2012 :

Items	Price		Quantity	
	2014	2017	2014	2017
P	20	25	10	12
Q	18	32	16	10
R	35	48	8	12
S	28	40	12	10

Sol :

Computation of the Laspeyre's Index number

Items	q_0	p_0	q_1	p_1	p_1q_0	p_0q_0
P	10	20	12	25	250	200
Q	16	18	10	32	512	288
R	8	35	12	48	384	280
S	12	28	10	40	480	336
Total	–	–	–	–	1626	1104

We have, $P_{01}(L) = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100 = \frac{1625}{1104} \times 100 = 147.28$

7. From the following data contruct Paasche's Index number for the year 2017 with the base 2015.

Items	Price		Quantity	
	Price	**Quantity**	**Price**	**Quantity**
P	**4**	**2**	**6**	**3**
Q	**3**	**5**	**2**	**1**
R	**8**	**2**	**4**	**6**
S	**3**	**4**	**5**	**8**

Sol :

Construction of the Paasche's Index number for 2017 base Year = 2015

Items	2015		2017		p_1q_1	p_0q_1
	q_0	p_0	p_1	q_1		
P	4	2	6	3	18	12
Q	3	5	2	1	2	3
R	8	2	4	6	24	48
S	3	4	5	8	40	24
Total	–	–	–	–	84	87

Paasche's Index number is given by

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100 = \frac{84}{87} \times 100 = 96.55$$

8. From the following data given below compute Fisher's ideal index for the year 2017

Commodity	2014		2017	
	Price	**Quantity**	**Price**	**Quantity**
P	**20**	**8**	**40**	**8**
Q	**50**	**10**	**60**	**5**
R	**40**	**15**	**50**	**15**
S	**10**	**20**	**20**	**25**

Sol :

Computation of the Laspeyre's Index number
for 2017 with 2014 as the base year

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
P	20	8	40	6	320	160	240	120
Q	50	10	60	5	600	500	300	250
R	40	15	50	15	750	600	750	600
S	10	20	20	25	400	200	500	250
Total	–	–	–	–	2070	1460	1790	1220

Fisher's ideal index is given by

$$P_{01}(F) = \sqrt{\frac{\sum p_1p_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$$

Substituting the respective values in the above formula we have,

$$P_{01}(F) = \sqrt{\frac{2070}{1460} \times \frac{1790}{1220}} \times 100$$

$$= \sqrt{1.4178 \times 1.4672} \times 100$$

$$= \sqrt{2.0802} \times 100 = 1.4423 \times 100$$

$$= \sqrt{2.0802} \times 100 = 1.4423 \times 100$$

$$= 144.23 \text{ approx.}$$

9. From the following data given below compute Fisher's ideal index for the year 2017

Items	2013		2017	
	Price	**Quantity**	**Price**	**Quantity**
P	**4**	**74**	**6**	**82**
Q	**10**	**125**	**8**	**140**
R	**14**	**40**	**12**	**33**

Sol :

Computation of the Laspeyre's Index number for 2017 with 2014 as the base year

Items	2013		2017					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
P	4	74	6	82	444	296	492	328
Q	10	125	8	140	1000	1250	1120	1400
R	14	40	12	33	480	560	396	462
					1924	2106	2008	2190

Marshall and Endgeworth's index number is given by

$$P_{01}(ME) = \frac{\sum p_1q_0 + \sum p_1q_1}{\sum p_0q_0 + \sum p_0q_1} \times 100 = \frac{1924 + 2008}{2106 + 2190} \times 100 = \frac{3932}{4296} \times 100$$

$$= 91.53 \text{ approx.}$$

Fisher's idel Index is given by

$$P_{01}(F) = \sqrt{\frac{\sum p_1p_0}{\sum p_0p_1} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100 = \sqrt{\frac{1924}{2106} \times \frac{2008}{2190}} \times 100$$

$$= \sqrt{0.9135 \times 0.9168} \times 100 = \sqrt{0.83749} \times 100 = 0.9151 \times 91.51$$

From the above results of the two indices, it is clear that the index number of Marshall and Edgeworth is a close approximation of the ideal index of Fisher.

10. Construct index numbers of price from the following data by applying :

(i) Laspeyres method

(ii) Paasche method

(iii) Fisher's ideal method, and

(iv) Marshall-Edgeworth method

Commodity	2010 Price	2010 Quantity	2011 Price	2011 Quantity
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

Sol :

Calculation of Various Indices

Commodity	2010 Price (P_0)	2010 Qty (Q)	2011 Price (P_1)	2011 Qty (q_1)	$p_1 q_0$	$p_0 q_0$	$p_1 q_1$	$p_0 q_1$
A	2	8	4	6	32	16	24	12
B	5	10	6	5	60	50	30	25
C	4	14	5	10	70	58	50	40
D	2	19	2	13	38	38	26	26
					200	160	130	103

(i) Laspeyres Method : $p_{01} = \frac{\Sigma\, p_1q_0}{\Sigma\, p_0q_0} \times 100$; where $\Sigma p_1q_0 = 200.\ \Sigma p_0q_0 = 160$

$$p_{01} = \frac{200}{160} \times 100 = 125$$

(ii) Paasche's Method : $p_{01} = \frac{\Sigma\, p_1q_1}{\Sigma\, p_0q_1} \times 100$; where $\Sigma p_1q_1 = 130.\ \Sigma p_0q_1 = 103$

$$p_{01} = \frac{130}{103} \times 100 = 126.21$$

(iii) Fisher's Ideal Method : $p_{01} = \sqrt{\frac{\Sigma\, p_1q_0}{\Sigma\, p_0q_0} + \frac{\Sigma\, p_1q_1}{\Sigma\, p_0q_1}} \times 100 = \sqrt{\frac{200}{160} + \frac{130}{103}} \times 100$

$$= \sqrt{1.578} \times 100 = 1.256 \times 100 = 125.6$$

(iv) Marshall-Edgeworth Method : $p_{01} = \dfrac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} \times 100 = \dfrac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1}$

$$= \frac{200 + 130}{160 + 103} \times 100 = \frac{330}{363} \times 100 = 125.47$$

11. From the following data calculate price index according to

(i) Laspeyre,

(ii) Paasche and

(iii) Marshall-Edgeworth methods.

Item	Base year		Current year	
	Price (`)	Expenditure (`)	Price (`)	Expenditure (`)
A	5	50	8	40
B	7	25	12	30
C	9	10	15	25
D	12	5	20	18

Sol : **(Jan.-21)**

$$\text{Qty} = \frac{\text{Expenditure}}{\text{Price}}$$

Base year

A $\dfrac{50}{5} = 10$

B $\dfrac{25}{7} = 3.6$

C $\dfrac{10}{9} = 1.1$

D $\dfrac{5}{12} = 0.42$

Base year

A $\dfrac{40}{8} = 5$

B $\dfrac{30}{12} = 2.5$

C $\dfrac{25}{15} = 1.7$

D $\dfrac{18}{20} = 0.9$

Item	p_0	q_0	p_1	q_1	p_1q_1	p_1q_0	p_0q_1	p_0q_0
A	5	10	8	5	40	80	25	50
B	7	3.6	12	2.5	30	43.2	17.5	25.2
C	9	1.1	15	1.7	25.5	16.5	15.3	9.9
D	12	0.42	20	0.9	18	8.4	10.8	5.04
					113.5	148.1	68.6	90.14

(i) Laspeyre method = $p_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$$= \frac{14.81}{90.14} \times 100 = 164.29$$

(ii) Passche method = $p_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$$= \frac{113.5}{68.6} \times 100 = 165.45$$

(iii) Marshall Edgeworth

$$p_{01} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

$$= \frac{148.1 + 113.5}{90.14 + 68.6} \times 100$$

$$= \frac{261.6}{158.74} \times 100 = 164.79$$

12. Compute Price Index number by using:

(i) Paasches and

(ii) Marshal and Edge worth methods.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
P	5	100	6	150
Q	4	80	5	100
R	2	60	5	72
S	12	30	9	33

Sol : **(June -19)**

Let base year price denoted as P_0.

Let base year quantity denoted as Q_0.

Let current year price denoted as P_1.

Let current year quantity denoted as Q_1.

Commodity	P_0	P_1	Q_0	Q_1	P_1Q_0	P_0Q_0	P_1Q_1	P_0Q_1
P	5	6	100	150	600	500	900	750
Q	4	5	80	100	400	320	500	400
R	2	5	60	72	300	120	360	144
S	12	9	30	33	270	360	297	396
					1570	1300	2057	1690

(i) Paasches Method

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{2057}{1690} \times 100$$

$$= 121.7159$$

(ii) Marshall and Edgeworth Method

$$P_{01} = \left[\frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1}\right] \times 100$$

$$= \left[\frac{1570 + 2057}{1300 + 1690}\right] \times 100$$

$$P_{01} = \frac{3627}{2990} \times 100$$

$$= 121.3043$$

13. From the following data calculate price index Number by using (i) Paasche's Method and (ii) Marshal Edgeworth Method.

Item	Base year		Current year	
	Price (Rs.)	Expenditure (Rs.)	Price (Rs.)	Expenditure (Rs.)
P	6	300	10	560
Q	2	200	2	240
R	4	240	6	360
S	10	300	12	288
T	3	120	8	240

Sol : **(June-18, Imp.)**

In this problem we are given the expenditure (e) and the prices (p) for different items. we have

$$\text{Expenditure} = \text{Price} \times \text{Quantity} \Rightarrow \text{Quantity} = \frac{\text{Expenditure}}{\text{Price}}$$

Base Year

$$P = \frac{300}{6} = 50$$

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

$$R = \frac{240}{4} = 60$$

$$S = \frac{300}{10} = 30$$

$$T = \frac{120}{3} = 40$$

Current Year

$$P = \frac{560}{10} = 56$$

$$Q = \frac{240}{2} = 120$$

$$R = \frac{360}{6} = 60$$

$$S = \frac{288}{12} = 24$$

$$T = \frac{240}{8} = 40$$

Item	p_0	q_0	p_1	q_1	p_0q_0	p_1q_1	p_0q_1	p_1q_0
P	6	50	10	56	300	560	336	500
Q	2	100	2	120	200	240	240	200
R	4	60	6	60	240	360	240	360
S	10	30	12	24	300	288	240	360
T	3	40	8	40	120	240	90	320
					1160	1688	1146	1740

(i) Paasche's Method

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{1658}{1146} \times 100 = 147.29$$

(ii) Marshal Edgeworth

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

$$= \frac{1740 + 1688}{1160 + 1146} \times 100$$

$$= \frac{3428}{2306} \times 100$$

$$= 148.06$$

2.5 Tests of Consistency of Index Number

2.5.1 Unit Test - Time Reversal Test - Factor Reversal Test - Circular Test

Q12. Explain the various tests of Consistency of Index Number.

Ans : **(Imp.)**

1. Unit Test

This test states that the formula of index number should be independent of the units in which the prices or quantities of various commodities (or items) are quoted. All the formulae, except the index number based on simple aggregate of prices (quantities) satisfy this test.

2. Time Reversal Test

This test was proposed by Prof. Irwin Fisher. According to 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 the base, or putting it another way, the index number reckoned forward should be reciprocal of the one reckoned backward.'

In simple terms, given two time periods I and II, if an index number is calculated for period II taking period I as base, its value should be the reciprocal value of the index number for period I taking period II as base. The index numbers for the purpose of the test, should be in decimal form and not in percentage form. In other words, $P_{0.1}$ and $P_{1.0}$ should not be multiplied with 100. Symbolically.

$$\boxed{P_{01} \times P_{1.0} = 1}$$

Where $P_{0.1}$, = Fishers index for period II taking period I as base and

$P_{0.1}$, = Fishers index for period I taking Period II as base

Time Reversal test is satisfied by Marshall-Edge worth, Fisher, Walsh. Kelly' s index numbers and also by simple Aggregative Index, simple geometric mean of price relatives and weighted average of price relatives. Laspeyre's and Paasche's index numbers do not satisfy the time reversal test.

3. Factor Reversal Test : This test was also proposed by proof Fisher. This test requires that the product of two index numbers, one measuring price taking quantities as base, and the other measuring quantities taking price as base, should be equal to the net increase in total value from one period to another. Let us illustrate the same with the help of an example.

If $P_{0,1}$, is price index number, $Q_{0,1}$, is quantity index number, the product of the two index numbers should be equal to the value index number V_{01}

$$\boxed{P_{0,1} \times Q_{0,1} = V_{0,1}}$$

Fishers index number satisfies the factor reversal test. No other method satisfies the factor reversal test.

4. Circular Test

The circular test was proposed by Weztergaard. It is an extension of the time-reversal test. It more than two time periods are considered, price index is calculated for each period with the previous year as base period. Lastly, the price index for the first year is calculated taking the last period as the base. The product of all the price index numbers should be equal to 1. Symbolically, if three time periods are considered, $P_{0,1} \times P_{1,2} \times P_{2,1} = 1$

Only simple geometric mean of price relatives method and Kelly's method satisfy the circular test.

PROBLEMS

14. Calculate Fisher's Ideal Index from the following data and prove that it satisfies both the time reversal and factor reversal tests :

Commodity	2010		2011	
	Price	Qty	Price	Qty
A	8	80	10	120
B	10	120	12	96
C	5	40	5	50
D	4	56	3	60
E	20	100	25	150

Sol :

Calculation of Fisher's Ideal Index

Commodity	P_0	q_0	P_1	q_1	$p_1 q_0$	$p_0 q_0$	$p_1 q_1$	$p_0 q_1$
A	8	10	10	12	100	80	120	96
B	10	12	12	8	144	120	96	80
C	5	8	5	10	40	40	50	50
D	4	14	3	20	42	56	60	80
E	20	5	25	6	125	100	150	120
					Σp_1q_0	Σp_0q_0	Σp_1q_1	Σp_1q_1
					451	396	476	426

$$p_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{451}{396} \times \frac{476}{426}} \times 100$$

$$\sqrt{1.2726} \times 100 = 1.125 \times 100 = 112.8$$

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

$$p_{01} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}} = \sqrt{\frac{426}{476} \times \frac{396}{451}}$$

$$P_{01} \times P_{10} = \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{476} \times \frac{396}{451}} = \sqrt{1} = 1$$

Hence, time reversal test is satisfied.

Factor Reversal Test : Factor reversal test is sastisfied when :

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

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

$$P_{01} \times P_{10} = \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{396} \times \frac{476}{451}} = \frac{476}{396}$$

This is also the value of $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$. Hence, the above data also satisfies the Factor Reversal Test

15. Construct a Fisher's Ideal Indeed from the following data and show that it satisfies time reversal and factor reversal test :

Commodity	1995		1996	
	P_0	q_0	P_1	q_1
A	10	40	12	45
B	11	50	11	52
C	14	30	17	30
D	8	28	10	29
E	12	15	13	20

Sol :

Construction of Fisher's Ideal Index

Items	P_0	q_0	P_1	q_1	$p_1 q_0$	$p_0 q_0$	$p_1 q_1$	$p_0 q_1$
A	10	40	12	45	480	400	540	450
B	11	50	11	52	550	550	572	572
C	14	30	17	30	510	420	510	420
D	8	28	10	29	280	224	290	232
E	12	15	13	20	195	180	260	240
					$\Sigma p_1 q_0$ = 2015	$\Sigma p_0 q_0$ = 1774	$\Sigma p_1 q_1$ = 2172	$\Sigma p_0 q_1$ = 1914

Fisher's Ideal Index : $p_{01} = \sqrt{\frac{\Sigma\ p_1 q_0}{\Sigma\ p_0 q_0} + \frac{\Sigma\ p_1 q_1}{\Sigma\ p_0 q_1}}$

$$= \sqrt{\frac{2015}{1774} \times \frac{2172}{1914}} \times 100 = 1.135 \times 100 = 113.5$$

Time Reversal Test : Time Reversal test is satisfied when :

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

$$P_{10}\sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{\frac{1914}{2172} \times \frac{1774}{2015}}$$

$$P_{01} \times P_{10} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{2172} \times \frac{1774}{2015}} = 1$$

Hence, time reversal test is satisfied by the given data

Factor Reversal Test : Factor reversal test is satisfied when :

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

$$p_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{1914}{1774} \times \frac{2172}{2015}}$$

$$p_{01} \times q_{01} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{1774} \times \frac{2172}{2015}} = \frac{2172}{1774}$$

$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$, is also equal to $\frac{2172}{1774}$. Hence factor reversal test is satisfied by the given data.

2.6 Base Shifting

Q13. What is Base Shifting ?

Ans :

For a variety of reasons, it frequently becomes necessary to change the reference base of an index number series from one time period to another without returning to the original raw data and recomputing the entire series. This change of reference base period is usually referred to as *"shifting the base".* There are two important reasons for shifting the base :

(i) The previous base has become too old and is almost useless for purposes of comparison. By shifting the base it is possible to state the series in terms of a more recent time period.

(ii) It may be desired to compare several index number series which have been computed on different base periods; particularly if the several series are to be shown on the same graph, it may be desirable for them to have the same base period. This may necessitate a shift in the base period.

When base period is to be changed, one possibility is to recompute all index numbers using the new base period. A simpler approximate method is to divide all index numbers for the various years corresponding to the old base period by the index number corresponding to the new base period, expressing the results as percentages. These results represent the new index numbers, the index number for the new base period being 100 per cent.

Mathematically speaking, this method is strictly applicable only if the index numbers satisfy the circular test.

$$\text{Index number with new base} = \frac{\text{Index of current year}}{\text{Index of new base year}} \times 100$$

PROBLEMS

16. Reconstruct the series of index numbers given below by shifting the base to 2010.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
Index No	100	120	132	140	150	164	180	208	220

Sol :

The given series of index numbers along with another series with 2010 = 100 is shown in Table. To illustrate, for 2004, we have (100/180) × 100 = 55.56

Shifting the Base Year

Year	Index No. 2004 = 100	Index No. 2010 = 100
2004	100	55.56
2005	120	66.67
2006	132	73.33
2007	140	77.78
2008	150	83.33
2009	164	91.11
2010	180	100.00
2011	208	115.56
2012	220	122.22

17. The index of 2010 is 100. It rises by 5% in 2011, falls 2% in 2012, rises 4% in 2013, rises 10% in 2014, falls 3% in 2015, and rises 8% in 2016. Find out the indices for the 7 years assuming that all the increases and decreases are the percentages of the respective proceeding years - Also recast the indices shifting base to 2014.

Sol : **(Jan.-21)**

Calculation of Index number for base year and change of base year 2014

Year	Old Index Number (Base Year 2010 = 100) $= \frac{100+\%}{100} \times$ Previous Year Index Number	New Index Number (New Base Year 2014 = 116.59) = $\frac{100}{\text{Value of New Base Year 2014}} \times$ old index number of year
2010	100 (Given)	$\frac{100}{116.59} \times 100 = 85.77$
2011	$\frac{100+2\%}{100} \times 100 = 104$	$\frac{100}{116.59} \times 104 = 89.20$
2012	$\frac{100-2\%}{100} \times 104 = 101.92$	$\frac{100}{116.59} \times 101.92 = 87.42$
2013	$\frac{100+4\%}{100} \times 101.92 = 105.99$	$\frac{100}{116.59} \times 105.99 = 90.91$

2014	$\frac{100+10\%}{100}$ × 105.99 = 116.59	$\frac{100}{116.59}\times 116.59$ = 99.99
2015	$\frac{100-3\%}{100}$ × 116.59 = 113.09	$\frac{100}{116.59}\times 113.09$ = 96.99
2016	$\frac{100+8\%}{100}$ × 113.09 = 122.14	$\frac{100}{116.59}\times 122.14$ = 104.76

18. The following are the indices (2007. Base)

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Indices	100	120	122	116	120	120	137	136	149	156	137

Shift the base to 2012 and recast the index numbers.

Sol. **(June-18, Imp.)**

$$\text{Current year F.B.I} = \frac{\text{Current tear C.B.I} \times \text{Previous year F.B.I}}{100}$$

The first year F.B.I being same as first year C.B.I; we obtain the F.B.I. numbers as given in table.

Conversion of C.B.I Numbers to F.B.I Numbers

Calculation of Index number for base year

Year	Incies	Fixed Base Index Number (Base 2012)
2007	100	$\frac{100}{120}\times 100 = 83.33$
2008	120	$\frac{120}{120}\times 100 = 100$
2009	122	$\frac{122}{120}\times 100 = 101.66$
2010	116	$\frac{116}{120}\times 100 = 96.66$
2011	120	$\frac{120}{120}\times 100 = 100$
2012	120	$\frac{100}{120}\times 100 = 100$
2013	137	$\frac{137}{120}\times 100 = 114.16$
2014	136	$\frac{136}{120}\times 100 = 113.33$
2015	149	$\frac{149}{120}\times 100 = 124.16$
2016	156	$\frac{156}{120}\times 100 = 130$
2017	157	$\frac{157}{120}\times 100 = 114.16$

2.7 Splicing of Index Numbers

Q14. What is Splicing?

Ans :

Combining two or more series of overlapping index numbers to obtain a single index number on a common base is called splicing of index numbers. Splicing of index numbers can be done only if the index numbers are constructed with the same items and have an overlapping year.

Splicing is generally done when an old index number with an old base is being discontinued and a new index with a new base is being started.

The process of splicing is very simple and is akin to that used in shifting the base. It is expressed in the form of a formula as follows :

$$\text{Spliced Index No. } \frac{\text{Index No. of current year} \times \text{Old Index of New Base Year}}{100}$$

PROBLEMS

19. The index A given was started in 1996 and continued up to 2006 in which year another indexed B was started. Splice the index B to index A so that a continuous series of index.

Year	Index A	Index B
1996	100	
1997	110	
1998	112	
2005	138	
2006	150	100
2007		120
2008		140
2009		130
2010		150

Sol :

Index B Spliced to Index A

Year	Index A	Index B	Index B spliced to index A 1982 as base
1996	100		
1997	110		
1998	112		
–			
–			
2005	138		

2006	150	100	$\frac{150}{100} \times 100 = 150$
2007		120	$\frac{150}{100} \times 120 = 180$
2008		140	$\frac{150}{100} \times 140 = 210$
2009		130	$\frac{150}{100} \times 130 = 195$
2010		150	$\frac{150}{100} \times 150 = 225$

The spliced index now refers to 1996 as base and we can make a continuous comparison of index numbers from 1996 onwards

In the above, it is also possible to splice the new index in such as manner that a comparison could be made with 2006 as base. This would be done by multiplying the old index by the ratio $\frac{100}{150}$. Thus the spliced index for 1996 would be $\frac{100}{150} \times 100 = 66.7$ for 1997, $\frac{100}{150} \times 110 = 73.3$ for 1998, $\frac{100}{150} \times 112 = 74.6$ etc. This process appears to be more useful because a recent year can be kept as a base. However, much would depend upon the object.

The continuation of series A or series B may be done by the equivalence of the index values in the overlapping year. Splicing of B to A is also called forward splicing while splicing of series A to B is called backward splicing.

20. Consider the following series of index numbers :

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
Series A: (2004=100)	100	120	150	180	220				
Series B : (2008=100)					100	110	150	160	175

a) Splice series A to series B

b) Splice series B to series A.

Sol :

Hence series A has base year as 2004 while series B has base year as 2008. The year overlapping in the two series is 2008 for which respective index value are 220 and 100. To splice A to B, we need to find the index values for the years 2004, 2005, etc. These can be obtained by multiplying the ratio 100/220 to the given index value of the year. For the year 2004, for example, we have

$$\text{Index value } = \frac{100 \times 100}{220} = 45.45$$

Splicing Index Number Series

Year	Series A	Series B	Splicing A to B	Splicing B to A
2004	100		45.45	100
2005	120		54.55	120
2006	150		68.18	150
2007	180		81.82	180
2008	220	100	100.00	220
2009		110	110.00	242
2010		150	150.00	330
2011		160	160.00	352
2012		175	175.00	385

2.8 Deflating of Index Numbers

Q15. What is Deflating ?

Ans :

The index numbers can be used to eliminate from a given series the effect of inflation over the long term. The income of a worker may be observed to increase over years on account of promotions, salary increments, etc. The question, however, is, does the purchasing power of his salary also change? This is because the prices also tend to rise over time. Now, if the rise in the prices is higher than the rise in his money income, then the purchasing power of his income, or the real income as it is called, would in fact decrease. If, however, the price increases are slower than the income increases, then the real income would rise. Thus, we calculate real income by adjusting the money income by appropriate price index. A similar calculation can be done to obtain the Gross Domestic Product (GDP) of a country in real terms for which GDP at market prices is adjusted for the price level changes. This process of eliminating the price effect from a given set of monetary values is termed deflating. The real values are obtained by dividing the monetary values by the price index value and multiplying the result by 100.

Q16. What are the various methods of Deflating?

Ans :

1. Purchasing power of money

This can be deflated by the following formula

$$\text{Purchasing power of money} = \frac{100}{\text{Price index}}$$

Thus, if the price rises by 25%, the price index becomes 125 and in that case, the purchasing power of every rupee would be $\frac{100}{125}$ = 0.80 ps. Or 80 ps. This means that a rupee in the current year is equal to 80 ps. in the base year.

2. Real wage, or income

This can be deflated by the following formula :

Real Wage $= \dfrac{100}{\text{Price Index}} \times$ Money Wage or $\boxed{\text{Real wage } = \dfrac{\text{Mone wage}}{\text{Price index}} \times 100}$

Here, price index should preferably be the consumer price index rather than the wholesale price index as the former reflects very well the change in the purchasing power of a wage earner.

Thus, if a worker earns ` 1500 during a year in which the price index stands at 150, the real value of his wage in comparison to the wage of the base year would be $\dfrac{1500}{150} \times 100 =$ ` 1,000. This means that his present wage of ` 1500 is equal to the wage of ` 1000 earned in the base year.

3. Real wage index, or Real income index

This can be deflated by the following formula :

Real Wage Index $= \dfrac{100}{\text{Price Index}} \times$ index of Money Wage

or Real wage index no. $= \dfrac{\text{Index of money wage}}{\text{Price index number}} \times 100$

or Again, $\boxed{\text{RWI} = \dfrac{\text{Real wage of the year}}{\text{Real wage of the base year}} \times 100}$

PROBLEMS

21. The table below shows the average wages in rupees per day of a group of industrial workers during the years 2005-2012. The consumer price indices for these years are also shown

a) Determine the real average wages of the workers furring the years

b) Calculate the index of real wages taking 2005=100

c) Represent the average actual and real wages graphically

Year	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Average Wage	**119**	**133**	**144**	**157**	**175**	**184**	**189**	**194**
CPI No	**100.0**	**107.6**	**106.6**	**107.6**	**116.2**	**118.9**	**119.8**	**120.2**

Sol :

The real wages may be calculated as follows :

Real wage $= \dfrac{\text{Average wage}}{\text{CPI No.}} = 100$

The illustrate, for the year 2006, we have

$$\text{Real wage} = \frac{133}{107.6} \times 100 = 123.61$$

Further, to calculate the real wage indices, we take the index for the year 2005 as 100. For each year, we have

$$\text{Real wage index} = \frac{\text{Average real wage for the year}}{\text{Average real wage for 2005}} \times 100$$

For example, for the year 2006, we have

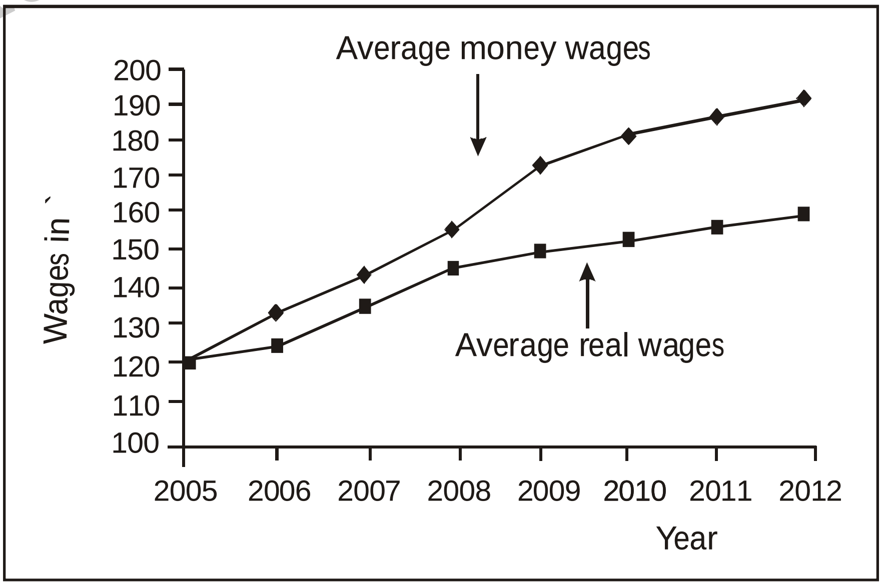
$$\text{Real wage index} = \frac{123.61}{119.00} \times 100 = 103.87$$

The results are given in Table

Calculation of Real Wages and Real Wages Indices

Year	Average Wage (`)	CPI No.	Average Real Wage(`)	Real Wage Index
2005	119	100.0	119.00	100.00
2006	133	107.6	123.61	103.87
2007	144	106.6	135.08	113.52
2008	157	107.6	145.91	122.61
2009	175	116.2	150.60	126.56
2010	184	118.9	154.75	130.04
2011	189	119.8	157.76	132.57
2012	194	120.2	161.40	135.63

The actual and real wages are shown plotted in figure. It is evident that while both of these are rising with time, the increase in real wages has been lower than the increase in money wages



22. The following table gives the annual income of worker and the general index numbers of price during 1988-1996. Prepare Index Number to show the changes in the real income of the teacher and comment on price increase.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Income Price	3600	4200	5000	5500	6000	6400	6800	7200	7500
Index No	100	120	145	160	250	320	450	530	600

Sol :

Index Number showing Changes in the Real Income of the Worker

Year	Income (Rs.)	Price Index No.	Real Income (Rs.)	Real Income Index No.
2001	3600	100	$\frac{3600}{100}\times 100 = 3600.00$	100.00
2002	4200	120	$\frac{4200}{120}\times 100 = 3500.00$	97.00
2003	5000	145	$\frac{5000}{145}\times 100 = 3448.27$	95.78
2004	5500	160	$\frac{5500}{160}\times 100 = 3437.50$	95.48
2005	6000	250	$\frac{6000}{250}\times 100 = 2400.00$	66.60
2006	6400	320	$\frac{6400}{320}\times 100 = 2000.00$	55.55
2007	6800	450	$\frac{6800}{450}\times 100 = 1511.11$	41.97
2008	7200	530	$\frac{7200}{530}\times 100 = 1358.49$	37.73
2009	7500	600	$\frac{7500}{600}\times 100 = 1250.00$	34.72

Q17. What are the limitations of Index Numbers?

Ans :

Despite the importance of the index numbers in studying the economic and commercial activities, and in measuring the relative changes in the price level as the economic barometers, they suffer from certain limitations for which they should be very carefully used and interpreted.

The following are some of the chief limitations among others :

1. They are only approximate indicators of the change of a phenomenon viz. price level, quantity level, cost of living, production activity etc. They never exactly represent the changes in die relative level of a phenomenon. This is because, the index numbers are constructed mostly on the sample data.

2. They are liable to be misused by a statistician with certain ulterior motive. If, purposively a wrong base year has been chosen, irrational weights have been assigned, inappropriate average has been used or irrelevant items have been included in the construction of an index number, the result would be highly misleading and fallacious.

3. They are prone to embrace errors at each, and every stage of construction viz :

   (i) Selection of the items and their numbers

   (ii) Obtaining the price quotations

   (iii) Selection of the base period

   (iv) Choice of the average

   (v) Assignment of weights

   (vi) Choice of the formula.

4. They are liable to misrepresent the true picture of a phenomenon, if the limited number of items chosen are not representative of the universe.

5. They are not capable of reflecting properly the relative changes in the quality level of the products which very much change in modem times.

Short Question and Answers

1. Explain the importance of index numbers.

Ans :

1. Economic Barometers

Index Numbers can be constructed for any phenomenon for which quantitative information is available. They capture the various changes taking place in the general economy and business activities. They provide a fair view of the general trade, the economic development and business activity of the country. Thus, they are aptly termed as 'Economic barometers'.

2. Study of Trend

Index Numbers study the relative changes in the level of a phenomenon over different periods of time. They are especially useful for study of general trend of the general trend for a group phenomenon in a time series data.

3. Policy Formulation

Index Numbers are indispensable for any organization in efficient planning and formulation of executive decisions. For example, the Dearness Allowance payable to employees is determined on the basis of Cost of living index numbers. Similarly, Psychiatrists use Intelligence Quotients to assess a child's intelligence in relation to his/her age, which further can be used in framing the education policy.

4. Deflation

Index Numbers can be deflated to find out the real picture pertaining to the phenomenon. For example, we can know if the real income of employees has been increasing by deflating the nominal wages with the help of index numbers.

5. Forecasting

Index Numbers provide valuable information that aids forecasting. For example an Index of sales, along with related indices such as cost of living index, helps in forecasting future demand and future sales of business.

2. Explain the various types of index numbers.

Ans :

An Index is an indicator. Index Numbers can be constructed for study of any phenomenon, if such phenomenon is capable of being quantified. In relation to data pertaining to business and economy, index numbers may be classified into the following.

1. Price Index Numbers

Price Index numbers measure the changes in prices. They are the most common indices. There are various methods by which the indices can be calculated. Most of the discussion in this chapter shall relate to price index numbers.

2. Quantity Index Numbers

Quantity Index Numbers study the changes in the volume of goods produced, consumed or transacted. The Index of Industrial Production is an example of a quantity index. However Quantity Index numbers are not very popular.

3. Value Index Numbers

Value Index numbers are rarely used. They study the change in total value, rather than mere change in prices of changes in quantities. They may have to be supplemented with price and quantity indices.

3. Define Index Numbers.

Ans :

Meaning

An Index, simply stated, is an indicator. It indicates the broad change in a given phenomenon. It is relative and indicates changes in the level of the given phenomenon in course of time, across geographical locations or in respect of any other characteristic. For example, the BSE Sensex (stands for Bombay Stock Exchange Sensitive Index) is an index of movement in stock prices. It tells us, very broadly, whether the prices of stocks of various companies have gone up or moved down on the Bombay Stock Exchange. Similarly, the Wholesale Price Index (WPI) gives us an indication of whether the general price level in the economy is going up or falling down. Similarly, Index numbers (or indices) can be constructed to measure any phenomenon, such as Industrial production, Number of road accidents, cost of Living or any other activity. Thus, Index Numbers are barometers measuring change in the level of a phenomenon.

Definitions

(i) **According to Croton and Cowden** Different Experts have defined Index Numbers in different words. Some of the definitions are stated as below.

(ii) **According to Speiges** "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

(iii) **According to Weldon** "An Index Number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographical location or any the characteristic."

(iv) **According to Edgeworth** "An Index Number is a statistical device for indicating the relative movements of data where measurement of actual movement is difficult or incapable of being made."

(v) **According to Bowley** "Index Number shows by its variations the changes in magnitude which is not susceptible either of accurate measurement in itself or of direct valuation in practice."

4. What are the characteristics of Index Numbers ?

Ans :

(i) **Index Numbers are Specialized Averages:** According to L.R. Connor, "In its simplest form an index number represents a special case of an average, generally a weighted average compiled from a sample of items judged to be representative of the whole."

In simple average the data treated have same units of measurement but index number average variables having different units of measurement. We may find an index number for a group of items including food, clothing, rent, cooking gas etc. which are all measured in different units.

(ii) **Index Numbers as Percentages and measure of Relative Change:** Index numbers are expressed in percentages but the word or symbol for percentage is never used. They measure relative changes in price over years with reference to some period called Base Period or Base Year. If the index number in 2000 is 150 with reference to 100 in 1995. It will mean that there is 50% increase in price in 2000 as compared to 1995. Since the index numbers are quantitative expressions of relative changes, they are expressed in numbers.

(iii) **Basis of Comparison :** Index numbers are used to make comparisons over different time periods with reference to base period.

(iv) **Universal Applications :** They are applied universally to ascertain different types of changes in different sectors of the economy. These changes may be with regard to prices, standard of living of the people, industrial or agricultural production, national or per capita income, exports or imports etc.

5. What is Laspeyre's index method ?

Ans :

This method takes the quantities of the commodities in the base period as the weight of that commodity for the purpose of calculating the index numbers. The following steps may be following.

Step 1:

Multiply the current year price (represented by p_0) with the quantities of the base year (q_0) for each commodity.

Step 2:

Add the numbers obtained in step 1. The resultant sum is represented as $\sum p_1q_0$

Step 3 :

Multiply the prices of base year (represented by p_0) with the quantities of the base year for each commodity.

Step 4 :

Add the numbers obtained in step 3. The resultant sum is represented as $\sum p_1q_0$

Step 5:

The index number as per Laspeyre's method

$$= P_{0.1} = \frac{\sum P_1q_0}{\sum P_0q_0} \times 100$$

---

6. What is Paasche's index method ?

Ans :

Paasche's Method is similar to Laspeyre's method. The only difference is in assignment of weights. As per this method quantities consumed of the commodities in the current year is taken as basis. The following steps need to be followed.

Step 1:

Multiply current year's prices (p_1) with current year's quantities (q_1)

Step 2 :

Add the numbers obtained in step (1). The resultant sum is $\sum p_1q_1$

Step 3 :

Multiply base year's Prices (p0) with current year's quantities (q_1)

Step 4 :

Add the numbers obtained in step (2). The resultant sum is $\sum p_0q_1$

Step 5 :

Index Number as per Paasche's method $= P_{0.1} = \dfrac{\sum p_1q_1}{\sum p_0q_1} \times 100$

Q7. Explain the comparision of Laspeyre's and Paasche's method ?

Ans :

S.No.	Laspeyre's Index Number	S.No.	Paasche's Index Number
1.	Here, quantity of the base year is assumed to be the quantity of the current year.	1.	Here, quantity of the current year is assumed to be the quantity of the base year.
2.	It has an upward bias i.e. the numerator of the index number is increased due to the assignment of higher weights fixed on the basis of the base year's quantities even though there might have been a fall in the quantity consumed during the current year due to rise, or fall in price and change in in tastes, habits and customs etc. in the current year.	2.	It has a downward bias i.e. the numerator of the index number is decreased due to the assignment of lower weights fixed on the basis of the current year's quantities even though the quantities in the current year might have fallen due to rise or fall in price, or change in habits of consumption.
3.	As the quantity of the base year are used as weights, the influence of price changes on quantities demanded do not get reflected in the index number.	3.	As the quantities of the current year are used as weights, the influence of price changes on quantities demanded get reflected in the index number.
4.	It measures changes in a fixed marked basketof goods and services as the same quantities are used in each period.	4.	It continually updates the quantities to the level of current consumption.
5.	Here, weights remain constant.	5.	Here, weights are determined every time an index number is constructed.

8. What is Marshall Edgeworth method ?

Ans :

In this method also both the current year as well as base year prices and quantities are considered. The formula for constructing the Index is :

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

on opening the brackets

$$P_{01} = \frac{\Sigma\, p_1q_0 + \Sigma\, p_1q_1}{\Sigma\, p_0q_0 + \Sigma\, p_0q_1} \times 100$$

It is a simple, readily constructed measure, giving a very close approximation to the results obtained by the ideal formula.

9. What is Fisher's Ideal Index ?

Ans :

Fisher's Ideal Index

This is the most popular amongst all weighted aggregative index numbers. It is obtained by calculating the Geometric Mean (G.M) of Laspeyre's and Paasche's index numbers. The formula for calculating fisher's ideal index is an under.

$$P_{0.1} = \left[\sqrt{\frac{\sum p_1q_0 \times \sum p_1q_1}{\sum p_0q_0 \times \sum p_0q_1}}\right] \times 100$$

Reasons for Fisher's Index being called an Ideal Index

Reasons

(i) It gives weightage to both current consumption and base year consumption.

(ii) It is free from upward or downward bias.

(iii) It satisfies both time reversal and factor reversal tests (to be discussed later) .

(iv) It is a Geometric mean of Laspeyre's index and Paasche's index

10. Time Reversal Test.

Ans :

This test was proposed by Prof. Irwin Fisher. According to 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 the base, or putting it another way, the index number reckoned forward should be reciprocal of the one reckoned backward.'

In simple terms, given two time periods I and II, if an index number is calculated for period II taking period I as base, its value should be the reciprocal value of the index number for period I taking period II as base. The index numbers for the purpose of the test, should be in decimal form and not in percentage form. In other words, $P_{0.1}$ and $P_{1.0}$ should not be multiplied with 100. Symbolically.

$$\boxed{P_{01} \times P_{1.0} = 1}$$

Where $P_{0.1}$, = Fishers index for period II taking period I as base and

$P_{0.1}$, = Fishers index for period I taking Period II as base

Time Reversal test is satisfied by Marshall-Edge worth, Fisher, Walsh. Kelly' s index numbers and also by simple Aggregative Index, simple geometric mean of price relatives and weighted average of price relatives. Laspeyre's and Paasche's index numbers do not satisfy the time reversal test.

11. Factor Reversal Test

Ans :

This test was also proposed by proof Fisher. This test requires that the product of two index numbers, one measuring price taking quantities as base, and the other measuring quantities taking price as base, should be equal to the net increase in total value from one period to another. Let us illustrate the same with the help of an example.

If $P_{0.1}$, is price index number, $Q_{0.1}$, is quantity index number, the product of the two index numbers should be equal to the value index number V_{01}

$$\boxed{P_{0.1} \times Q_{0.1} = V_{0.1}}$$

Fishers index number satisfies the factor reversal test. No other method satisfies the factor reversal test.

12. Circular Test

Ans :

The circular test was proposed by Weztergaard. It is an extension of the time-reversal test. It more than two time periods are considered, price index is calculated for each period with the previous year as base period. Lastly, the price index for the first year is calculated taking the last period as the base. The product of all the price index numbers should be equal to 1. Symbolically, if three time periods are considered, $P_{0.1} \times P_{1,2} \times P_{2,1} = 1$

Only simple geometric mean of price relatives method and Kelly's method satisfy the circular test.

13. What is Base Shifting ?

Ans :

For a variety of reasons, it frequently becomes necessary to change the reference base of an index number series from one time period to another without returning to the original raw data and recomputing the entire series. This change of reference base period is usually referred to as *"shifting the base".* There are two important reasons for shifting the base :

(i) The previous base has become too old and is almost useless for purposes of comparison. By shifting the base it is possible to state the series in terms of a more recent time period.

(ii) It may be desired to compare several index number series which have been computed on different base periods; particularly if the several series are to be shown on the same graph, it may be desirable for them to have the same base period. This may necessitate a shift in the base period.

When base period is to be changed, one possibility is to recompute all index numbers using the new base period. A simpler approximate method is to divide all index numbers for the various years corresponding to the old base period by the index number corresponding to the new base period, expressing the results as percentages. These results represent the new index numbers, the index number for the new base period being 100 per cent.

Mathematically speaking, this method is strictly applicable only if the index numbers satisfy the circular test.

$$\text{Index number with new base} = \frac{\text{Index of current year}}{\text{Index of new base year}} \times 100$$

14. What is Splicing?

Ans :

Combining two or more series of overlapping index numbers to obtain a single index number on a common base is called splicing of index numbers. Splicing of index numbers can be done only if the index numbers are constructed with the same items and have an overlapping year.

Splicing is generally done when an old index number with an old base is being discontinued and a new index with a new base is being started.

The process of splicing is very simple and is akin to that used in shifting the base. It is expressed in the form of a formula as follows :

$$\text{Spliced Index No.} \ \frac{\text{Index No. of current year} \times \text{Old Index of New Base Year}}{100}$$

15. What is Deflating ?

Ans :

The index numbers can be used to eliminate from a given series the effect of inflation over the long term. The income of a worker may be observed to increase over years on account of promotions, salary increments, etc. The question, however, is, does the purchasing power of his salary also change? This is because the prices also tend to rise over time. Now, if the rise in the prices is higher than the rise in his money income, then the purchasing power of his income, or the real income as it is called, would in fact decrease. If, however, the price increases are slower than the income increases, then the real income would rise. Thus, we calculate real income by adjusting the money income by appropriate price index. A similar calculation can be done to obtain the Gross Domestic Product (GDP) of a country in real terms for which GDP at market prices is adjusted for the price level changes. This process of eliminating the price effect from a given set of monetary values is termed deflating. The real values are obtained by dividing the monetary values by the price index value and multiplying the result by 100.

Exercise Problems

1. From the following data, calculate index numbers by simple aggregative method.

	A	B	C	D
Price in 1990 (Rs.)	162	256	257	132
Price in 1991 (Rs.)	171	164	189	145

[Ans : 83.90]

2. From the following data, calculate index numbers by simple aggregative method for the year 2001.

	A	B	C	D	E
Price in 2000 (Rs.)	150	160	175	190	200
Price in 2001 (Rs.)	155	172	170	205	190

[Ans : 101.94]

3. Find by the Arithmetic Mean method, the index number from the following data.

Commodity	Base Price (Rs.)	Current Price (Rs.)
Rice	30	35
Wheat	22	25
Fish	54	64
Potato	20	25
Coal	15	18

[Ans : 118.76]

4. Calculate the index number for the year 1989 with 1980 as base from the following data using weighted average of price relatives.

Commodity	Weights	1980 prices (Rs.)	1989 Prices (Rs.)
A	22	2.50	6.20
B	48	3.30	4.40
C	17	6.25	12.75
D	13	0.65	0.90

[Ans ; Using AM : 171.23, GM : 165.1]

5. The Price quotations of four different commodities for 1981 and 1985 are given below. Calculate the index number for 1985 with 1980 as base by using (i) simple average of price relatives (ii) weighted average of Price relatives.

Commodity	Weight (Rs.)	1985 Price (Rs.)	1980 Price
A	5	4.50	2.00
B	7	3.20	2.50
C	6	4.50	3.00
D	2	1.80	1.00

[Ans : (i) 170.75 (ii) 164.05]

6. Compute by Fishers index formula, the quantity index number from the data given below.

Commodity	Base Year		Current Year	
	Price (`)	Total Value (`)	Price (`)	Total Value (`)
A	10	100	8	96
B	16	96	14	98
C	12	36	10	40

[Ans : 120.65]

7. The following figures relate to the prices and quantities of certain commodities. Using Kelly's method, construct an Index number:

Commodity	Quantity Consumed	Price in 1980 (`)	Price in 1985 (`)
A	50	32	40
B	35	30	42
C	55	16	24
D	45	40	52
E	15	35	42

[Ans: 132.53]

Choose the Correct Answer

1. For comparing yearly changes in price level, the suitable index to be used is : [c]

   (a) F.B.I. with average price as base (b) F.B.I.

   (c) C.B.I. (d) None of these

2. The weights used in Laspeyre's price index are denoted as : [a]

   (a) q_0 (b) q_1

   (c) P_0 (d) P_1

3. The weights used in Laspeyre's quantity index are denoted as : [c]

   (a) q_0 (b) q_x

   (c) p_n (d) p_1

4. The weights used in Paasche's price index are denoted as : [b]

   (a) q_0 (b) q_1

   (c) p_0 (d) p_1

5. The weights used in Paasche's quantity index are denoted as : [d]

   (a) q_0 (b) q_1

   (c) p_0 (d) p_1

6. Weighted aggregative index formula using base year quantities as base is called : [a]

   (a) Laspeyre's price index (b) Paasche's price index

   (c) Bowley's price index (d) Fisher's price index

7. Weighted aggregative index formula using formula using the average of base year and current year's quantities as weights is called : [c]

   (a) Laspeyre's price index (b) Fisher's price index

   (c) Marshall-Edgeworth's index (d) Bowley's index

8. The geometric mean of Laspeyre's and Paasche's indices is : [a]

   (a) Fisher's ideal index (b) Bowley's index

   (c) Marshall and Edgeworth's index (d) None of these

9. Weighted average of relatives if base year value is taken as weights gives : [b]

(a) Fisher's index (b) Laspeyre's index

(c) Paasche's index (d) Bowley's index

10. The formula for simple average of price relative is : [a]

(a) $\frac{1}{n}\sum\frac{p_1}{p_0}\times 100$ (b) $\frac{\sum p_1}{\sum p_0}\times 100$

(c) $\frac{1}{n}\sum\frac{q_1}{q_0}\times 100$ (d) None of these

Fill in the blanks

1. Consumer price index number is measures _______.
2. If the price index increases by 20% the product A, which at present is Rs. 10, will _______ (increase by Rs. 2, increase by Rs. 12)
3. Index numbers are _______.
4. Index of industrial production is a _______.
5. Index numbers indicate _______.
6. Index numbers are known as _______.
7. Index numbers measure changes over lime in magnitudes which are not capable of _______.
8. An index number is a special type of _______.
9. Index numbers are expressed in _______.
10. Index numbers can be used for _______.
11. To measure changes in the price level for a group of people _______ index prepared.
12. Index numbers are called _______ of economic changes.
13. _______ test is satisfied by both Fisher's and Kelley's formulae.
14. The base period should be a _______ period
15. Fisher's index is _______ mean of Laspeyre and Passche index number.
16. Quantity index number reflects _______ changes from one period to another
17. _______ is the most suitable average for constructing index numbers.

ANSWERS

1. Measures changes in retail price
2. Rs. 2
3. Specialised averages
4. Quantity index,
5. Relative changes.
6. Economic barometers
7. Direct measurement

8. Average
9. Percentages
10. Forecasting
11. Cost of living
12. Barometer
13. Factor reversal
14. Normal
15. geometric
16. Quantity
17. Geometric mean

UNIT III	**TIME SERIES :** Introduction - Components - Methods-Semi Averages - Moving Averages - Least Square Method - Deseasonalisation of Data - Uses and Limitations of Time Series.

3.1 Time Series

3.1.1 Introduction

Q1. Define time series.

(OR)

What is time series.

Ans : **(Imp.)**

Meaning

A time series is a statistical data that are collected, observed or recorded at regular intervals of time. The term time series applies, for example, to the data recorded periodically showing the total annual sales of retail stores, the total quarterly value of construction contracts awarded, the total amount of unfilled orders in durable goods industries at the end of each month, weekly earnings of workers in an industrial town, hourly temperature in a particular city.

Definitions

Some of the important definitions of time series, given by different experts are as under:

i) **According to Morris Hamburg,** "A time series is a set of statistical observations arranged in chronological order."

ii) **According to Patterson,** "A time series consists of statistical data which are collected, recorded observed over successive increments."

iii) **According to Ya-Lun-Chou,** "A time series may be defined as a collection of magnitudes belonging to different time periods, of some variable or composite of variables, such as production of steel, per capita income, gross national product, price of tobacco, or index of industrial production."

iv) **According to Wessel and Wellet,** "When quantitative data are arranged in the order of their occurrence, the resulting statistical series is called a time series."

v) **According to Spiegel,** "A time series is a set of observations taken at specified times, usually at 'equal intervals'. Mathematically, a time series is defined by the values Y_1 Y_2 ... of a variable Y (temperature, closing price of a share, etc.) at times t_1 , t_2 .... Thus Y is a function of t_1 symbolized by Y= F(t)."

vi) **According to Cecil H. Mayers,** "A time series may be defined as a sequence of repeated measurement of a variables made periodically through time".

It is clear from the above definitions that time series consist of data arranged chronologically. Thus if we record the data relating to population, per capita income, prices, production, etc., for the last 5, 10, 15, 20 years or some other time period, the series so emerging would be called time series.

It should be noted that the term 'time series' is usually used with reference to economic data and the economists are largely responsible for the development of the techniques of time series analysis. However, the term 'time series' can apply to all other phenomena that are related to time such as the number of accidents occurring in a day, the variation in the temperature of a patient during a certain period, number of marriages taking place during a certain period, etc.

Q2. What are the characteristics of time series ?

Ans : **(Imp.)**

The essentials of Time Series are

i) It must consist of a set of values that are homogeneous. For example, production data for a year and sales data for the next year will not be a time series.

ii) The values must be with reference to time. In other words, in a time series, we have at least 2 variables, with one variable necessarily being time.

iii) The data must be available for a reasonably long period of time.

iv) The gaps between various time values should as far as possible be equal.

v) The values of the second variable should be related to time. For example, the number of people being hired by the BPO industry can be tracked in relation to time. However, if we are talking about average height of students in a class, this data may not have a significant relationship with time and may not constitute a time series.

Q3. What are the objectives of time series ?

Ans :

While it is true that past performance does not necessarily guarantee future results, the quality of forecasts that management can make is strongly related to the information that can be extracted and used from past data. Thus, the objective of time series analysis is to interpret the changes in a given variable with reference to the given situation and attempt to anticipate the future course of events. Analysis of Time Series is done with the following objectives:

i) To evaluate past performance in respect of a particular variable.

ii) To make future forecasts in respect of the particular variable.

iii) To chart short term and long term strategies of the business in respect of the particular variable.

3.2 Components of Time Series

Q4. What are the Components of Time Series?

Ans : **(Jan.-21, Imp.)**

The changes in the values of a variable related to time can be the result of a large Variety of factors such as change in the tastes and habits of people, change in population, change in cost of production, change in income of people, change in relationship between countries, addition or elimination of competitors, change in climatic conditions, change in Government policy etc. The value of a variable changes due to interaction of such factors. The various factors affecting the values of the given variable may be broadly classified into four categories, commonly known as Components of Time Series.

Each of these components explains some distinct characteristics of change, long term or short term, regular or irregular or non-repetitive. The various components of Time Series can be broadly classified as under:

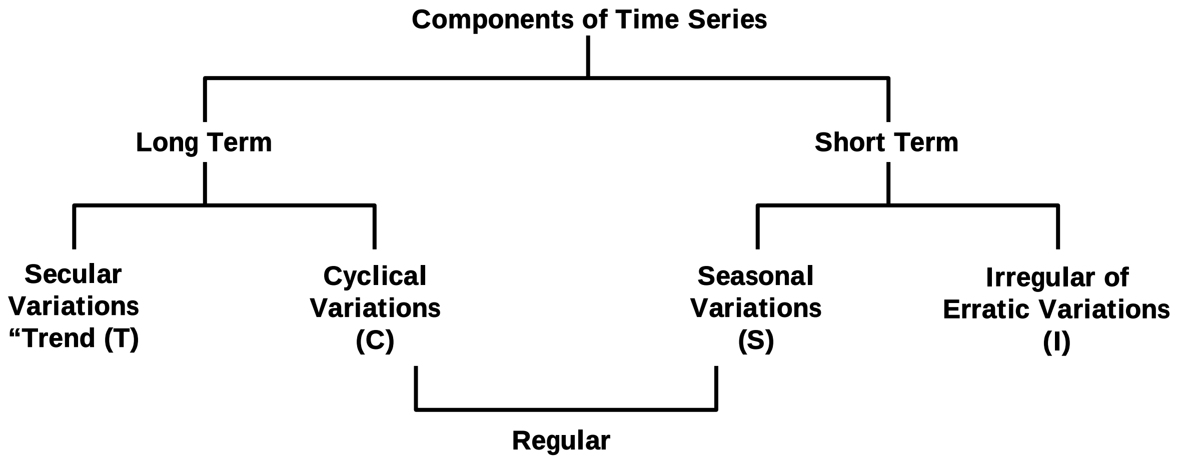


Fig.: Components of Time Series

Thus, the components of Time Series are:

i) Secular Trend or Long Term Movement

ii) Seasonal Variations

iii) Cyclical Variations and

iv) Irregular, Random or Erratic Variations

i) Secular Trend or Long Term Movement

Secular Trend is the basic tendency of a series to grow, decline, remain constant or fluctuate, over a long period of time. The concept of trend does not include short term changes but is concerned with steady movement over a long period. The long term trend movement is the result of forces that experience change very gradually and continuously over a long period of time. They operate in an evolutionary manner and do not reflect sudden changes. For example, the number of people travelling by Airways has gradually and continuously gone up. The number of infant deaths per thousand children bom is a steadily declining trend. This gradual and continuous movement of trend can be attributed to various factors such as increased investment in infrastructure, opening up of the economy through economic reforms, advances in technology, change in demographic profile etc.

The long term trend helps us in determining the direction of change. The growth factor can be estimated with a fair degree of accuracy. It provides indication of what is ahead for a given series. Elimination of trend from the original data helps in understanding the other elements that influence data in the short run.

ii) Seasonal Variations

Seasonal variations involve patterns of change within a year that tend to get repeated year after year. The trend is fluctuating and repetitive as it peaks and bottoms out at about the same time of the year, every year. Seasonal variations can be detected only if the data is recorded in smaller units of time such as weekly, monthly or quarterly. There are no seasonal variations in a time series where only annual figures are available. These are the result of such factors that uniformly rise and fall in magnitude. For example, the sale of umbrellas peaks in the months of June and July every year. This is on account of the onset of monsoons. The prices of agricultural commodities fall at the time of harvest. The passenger traffic increases substantially in the summer vacations. In all these examples, the movement of the trend is for a short period of time (season). The same movements are repeated in the coming years. Hence, it is easy to forecast the future.

Characteristics: Based on the above discussion, we can list the following characteristics of seasonal variations.

a) They repeat themselves periodically in less than one year's time.

b) These are results of factors that uniformly and regularly rise and fall in magnitude.

c) These variations are periodic and regular. They can be predicted without much difficulty.

Causes: Seasonal Variations can be on account of natural forces or due to manmade conventions. For example, the trend in umbrellas can be attributed to nature, while increase in passenger traffic in summer is manmade. However in either situation, it is very important to study the seasonal variations. A proper understanding of seasonal variations leads to better planning of future operations. If a seasonal upswing in sales is misinterpreted as a genuine increase in demand for the product, the entire operations of the business will get adversely affected.

iii) Cyclical Variations

The cyclical variations in a series are the recurrent variations whose duration is more than one year. Cyclical variations are regular but not uniformly periodic. It is relatively more difficult to predict the future direction of the series. One complete period that normally lasts from seven to nine years, is termed as a cycle. The most common example of Cyclical Variations is the business cycle. The business cycle passes through four phases of Boom, Recession, Depression and Recovery and it may take more than 10 years to complete the cycle.

Characteristics: The characteristics of a cyclical variation are

(a) There are Oscillating movements above and below the secular tread line.

(b) The fluctuations occur over a period greater than one year.

(c) There is no uniformity in the period of recurrence of movement pattern. One cycle may get completed in 3 years while the next cycle for the same date may takes 8 years.

(d) Cyclical fluctuations are more difficult to measure.

Utility

(a) Knowledge of cyclical variations helps users to understand that they will not be perennially in the same state of boom or depression. They realize the need or different strategies at different times.

(b) Study of cyclical variations helps businesses to predict the turning points ahead of time.

(c) Understanding of cyclical variations is helpful in formulation of policies such as diversification aimed at stabilizing business fluctuations.

Challenges

(a) There is no uniformity in the period of recurrence of movement pattern.

(b) They ,are mixed with Irregular Variations and it is difficult to separate the two.

iv) Irregular, Random (or) Erratic Variations

In many situations, the value of a variable may be completely unpredictable, changing in a random manner. These fluctuations are the result of such unforeseen and unpredictable forces that operate in absolutely erratic and irregular manner. There is no definite pattern and there is no regular period of time of their occurrence.

They are normally short-term variations caused by non-recurring factors such as floods, famines, revolution (such as green revolution) etc. However, in rare instances, the effects of these random variations are such that they may lead to new cyclical or seasonal movements. Irregular variations are also known as 'episodic' variations and include all variations other than those accounted for by trend, seasonal and cyclical variations.

Need for Studying Irregular Variations

There are two reasons for recognizing irregular movements:

(a) To suggest that, on occasions, it may be possible to explain certain movements in the data as due to specific causes and to simplify further analysis.

(b) To emphasize the fact that predictions of economic conditions are always subject to a degree of error owing to the unpredictable erratic influences.

Adjustments to be made before Analyzing Time Series

Before analyzing a time series, it is necessary to make certain adjustments in the available data. The adjustments are necessary to bring about homogeneity in data in the following points:

i) Time Variations: When data is available on monthly basis, effects of time variation need adjustment. All months of the year do not have the same number. Within the month, the number of working days may be differer'. For example, September has 30 days and October has 31 days. Also, while August may also have 31 days, the number of holidays (including Sundays) may be more in October. Thus, if we are analyzing monthly production data and if the number of units produced in a day is identical on each day of the month, different months with different working days would show variation in the production figure, although there is no real variation. Thus, it is necessary to adjust for Calendar Variation before we start the analysis.

ii) Price Changes: Adjustment for price changes become necessary wherever we have a current value series and we are interested in real value changes. In such a case, the current values have to be deflated by the ratio of current prices to base year paces to arrive at real values. Such adjustments are also necessary when we have information about value but we need to study the trend in quantity changes.

iii) Population Changes: Adjustments for population change become necessary when a variable is affected by change in population. If we are studying national income figures or figures of output, such adjustment becomes necessary. In such cases, it is preferable to study the per capita income or per capita output figures, which can be arrived at by dividing the income or output by the number of persons concerned.

iv) Comparability of the data: The data which is being analyzed should be comparable. However, data relating to past may not be strictly homogenous with data in the present. This can be on account of various factors such as change in definitions of the term, change in the size of the sample, restructuring on account of mergers, etc. For example, sales data in some years could be gross sales while, owing to change in policy, the sales data for next few years can be Net Sales. Similarly, if the state of Madhya Pradesh is split and a new state of Chattisgarh is created, data of Madhya Pradesh before and after the creation of Chattisgarh will not be comparable.

Q5. What are the various methods of seasonal variations? Explain simple average.

Ans :

The following are the various methods of seasonal variations are:

Methods of seasonal variations :

a) Method of Simple Averages (Weekly, Monthly or Quarterly).

b) Ratio-to-Trend Method.

c) Ratio-to-Moving Average Method.

d) Link Relative Method.

a) Simple Averages

This is the simplest method of obtaining a seasonal index. The following steps are necessary for calculating the index:

(i) Arrange the unadjusted data by years and months (or, quarters if quarterly data are given).

(ii) Find the totals of January, February, etc.

(iii) Divide each total by the number of years for which data are given. For example, if we are given monthly data for five years then, we shall first obtain total for each month for five years and divide each total by 5 to obtain an average.

(iv) Obtain an average of monthly averages by dividing the total of monthly averages by 12.

(v) Taking the average of monthly averages as 100, compute the percentages of various monthly averages as follows :

$$\text{Seasonal Index for January} = \frac{\text{Monthly average for January}}{\text{Average of monthly averages}} \times 100$$

Problems

1. Assuming that trend is absent, determine if there is any seasonality in the data given below :

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2006	3.7	4.1	3.3	3.5
2007	3.7	3.9	3.6	3.6
2008	4.0	4.1	3.3	3.1
2009	3.3	4.4	4.0	4.0

What are the seasonal indices for various quarters ?

Sol :

Computation of Seasonal Indices				
Year	**1st Quarter**	**2nd Quarter**	**3rd Quarter**	**4th Quarter**
2006	3.7	4.1	3.3	3.5
2007	3.7	3.9	3.6	3.6
2008	4.0	4.1	3.3	3.1
2009	3.3	4.4	4.0	4.0
Total	14.7	16.5	14.2	14.2
Average	3.675	4.125	3.55	3.55
Seasonal index	98.66	110.74	95.30	95.30

Notes for calculating Seasonal index

$$\text{The average of averages} = \frac{3.675 + 4.125 + 3.55 + 3.55}{4} = \frac{14.9}{4} = 3.725$$

$$\text{Seasonal index for the first quarter} = \frac{3.675}{3.725} \times 100 = 98.66$$

$$\text{Seasonal index for the second quarter} = \frac{4.125}{3.725} \times 100 = 110.74$$

$$\text{Seasonal index for the third and fourth quarter} = \frac{3.55}{3.725} \times 100 = 95.30$$

2. Calculate the seasonal index for the following data using the simple average method :

Year	Quarters			
	I	II	III	IV
2002	35	39	34	36
2003	35	41	37	48
2004	35	39	37	40
2005	40	46	38	45
2006	41	44	42	45
2007	42	46	43	47

Sol :

Year	Quarters			
	I	II	III	IV
1988	35	39	34	36
1989	35	41	37	48
1990	35	39	37	40
1991	40	46	38	45
1992	41	44	42	45
1993	42	46	43	47
Quarterly Total	228	255	231	261

Quarterly averages; $\frac{228}{6} = 38$; $\frac{225}{6}$ = 42.5; $\frac{231}{6}$ = 38.5; $\frac{261}{6}$ = 43.5

Average of quarterly averages = $\frac{38 + 42.5 + 38.5 + 43.5}{4} = \frac{162.5}{4} = 40.63$

Seasonal index for first quarter = $\frac{\text{Quarterly average for first quarter}}{\text{Averge of quarterly averages}} \times 100$

$= \frac{38}{40.63} \times 100 = 93.53$

Seasonal index for second quarter = $\frac{42.5}{40.63} \times 100 = 104.60$

Seasonal index for third quarter = $\frac{38.5}{40.63} \times 100 = 94.75$

Seasonal index for fourth quarter = $\frac{43.5}{40.63} \times 100 = 107.06$

3. Compute the seasonal index numbers applying simple average method for the following data :

Year	Summer	Monsoon	Autumn	Winter
2001	112	110	120	115
2002	80	145	105	90
2003	95	100	140	80
2004	110	90	130	110
2005	85	110	110	90
2006	92	120	100	85

Sol :

Year	Summer	Monsoon	Autumn	Winter
2001	112	110	120	115
2002	80	145	105	90
2003	95	100	140	80
2004	110	90	130	110
2005	85	110	110	90
2006	92	120	100	85
N = 6 Total	574	675	705	570
Seasonal Average	95.97	112.50	117.50	95.00
Seasonal Index Nos.	90.97	106.97	111.73	90.33

General average = Average of seasonal averages

$$= \frac{95.67 + 112.50 + 117.50 + 95.00}{4}$$

$$= \frac{420.67}{4} = 105.167$$

$$\text{Seasonal index} = \frac{\text{Seasonal average}}{\text{General average}} \times 100$$

$$\text{Seasonal index for summer} = \frac{95.67}{105.167} \times 100 = 90.97\ ;$$

$$\text{Seasonal index for monsoon} = \frac{112.50}{105.167} \times 100 = 106.97;$$

$$\text{Seasonal index for autumn} = \frac{117.50}{105.167} \times 100 = 111.73;$$

$$\text{Seasonal index for winter} = \frac{95}{105.167} \times 100 = 90.33$$

Q6. Explain briefly about ratio to trend method.

Ans :

This method of calculating a seasonal index (also known as the percent- age-to-trend method) is relatively simple and yet an improvement over the method of simple averages explained in the preceding section. This method assumes that seasonal variation for a given month is constant fraction of trend. The ratio-to-trend method presumably isolates the seasonal factor in the following manner. Trend is eliminated when the ratios are computed, in effect

$$\frac{T \times S \times C \times I}{T} = S \times C \times I.$$

Random elements are supposed to disappear when the ratios are averaged. A careful selection of the period of years used in the computation is expected to cause die influences of prosperity or depression to offset each other and thus reipove the cycle. For series that are not subject to pronounced cyclical or random influences grid for which trend can be computed accurately, this method may suffice.

Limitations of the Ratio-to-Trend Method

Merits

- Compared with the method of monthly averages this method is certainly a more logical procedure for measuring seasonal variations. It has an advantage over the moving average procedure too. For it has a ratio-to-trend value for each month for which data are available. Thus there is no loss of data as occurs in the case of moving averages. This is a distinct advantage especially, when the period covered by the time series is very short.
- It is simple to compute and easy to understand.

Limitations

The main defect of the ratio-to-trend method is that if there are pronounced cyclical swings in the series, the trend - whether a straight line or a curve-can never follow the actual data as closely as a 12-month moving average does. In consequence as seasonal index computed by the ratio to moving average method may be less biased than the one calculated by the ratio-to-trend method.

Steps

(i) Find the trend values with the help of method of least squares.

(ii) Divide the given original data (quarterly or monthly) by corresponding trend value and multiply this by 100. The values so obtained are free from trend.

(iii) Find the quarterly or monthly (as the case may be) average of trend eliminated values. The average value is free from irregular and cyclical movements or variations.

(iv) Add the quarterly (or monthly) averages say (S), divide 400 (or 1200) by S. This is called constant factors.

(v) Multiply each quarterly (or monthly) average obtained is step (iii) by the constant factor obtained in step (iv).

PROBLEMS

4. For the given data below compute seasonal variations using ratio to trend method.

Year	Quarters			
	I	II	III	IV
2003	60	80	72	68
2004	68	104	100	88
2005	80	116	108	96
2006	108	152	136	124
2007	160	184	172	164

Sol:

First determine the trend on yearly basis and convert it into quarterly data.

Year X	Yearly Total	Quarterly Average (Y)	Deviations from Mid-year i.e., 2005 x = X – 1998	xY	x^2	Y_c
2003	280	70	-2	-140	4	64
2004	360	90	-1	-90	1	88
2005	400	100	0	0	0	112
2006	520	130	1	130	1	136
2007	680	170	2	340	4	160
N = 5		$\Sigma Y = 560$	0	$\Sigma xY = 240$	$\Sigma x^2 = 10$	

$$a = \frac{\Sigma Y}{N} = \frac{560}{5} = 112$$

$$b = \frac{\Sigma xY}{\Sigma x^2} = \frac{240}{10} = 24$$

$$y_c = 112 + 24x = 112 + 24\,(X - 1998)$$

Yearly increment = 24

$$\text{Quarterly increment} = \frac{24}{4} = 6$$

Calculation of quarterly trend values:

Consider 1996, trend value for middle quarter, i.e., half of 2nd ad half of 3rd is 64. Quarterly increment is 6. So the trend value of 2nd quarter is $64 - \frac{6}{2} = 61$ and for 3rd quarter is $64 + \frac{6}{2} = 67$. Trend value for the 1st quarter is 61 – 6 = 55 and of the 4th quarter is 67 + 6 = 73.

Quarterly Trend Values

Year	I	II	III	IV
2003	55	61	67	73
2004	79	85	91	97
2005	103	109	115	121
2006	127	133	139	145
2007	151	157	163	169

The given values of the time series will now be expressed as percentages of the corresponding trend values given above. These are trend eliminated values.

Year	I	II	III	IV
2003	109.09	131.15	107.46	93.15
2004	86.08	122.35	109.89	90.72
2005	77.67	106.42	93.91	79.34
2006	85.04	114.29	97.84	85.52
2007	105.96	117.20	105.52	97.04
Total	463.84	591.41	514.62	445.77
Quarterly Average	92.77	118.28	102.92	89.15
Seasonal Average	92.05	117.36	102.12	84.47

$\frac{60}{55} \times 100 = 109.09, \frac{80}{61} \times 100 = 131.15$, etc.

Sum of the quarterly averages = $\frac{92.77 + 118.28 + 102.92 + 89.15}{4} = 403.12$

Constant factor = $\frac{400}{403.12} = 0.99226$

Seasonal index for 1st quarter = 92.77 × 0.99226 = 92.05

Seasonal index for 2nd quarter = 118.28 × 0.99226 = 117.36 and so on.

Q7. Explain briefly about ratio to moving average method.

Ans :

This method is an improvement over ratio to trend method because it tries to eliminate the cyclic variations which are mixed with the seasonal indices in ratio to trend method. In this method moving average trend is used in place of least square trend.

The following steps are involved :

(i) Take centered 12 monthly (or four quarterly) moving average. Since variations occur after 12 months for a monthly data, a 12-month moving average will completely eliminate the seasonal fluctuations if they are regular and constant, Even irregular variations will be eliminated due to averaging. Thus moving average will contain only trend (T) and cyclic variations (C).

(ii) Express the original data for each month as a percentage of the centered 12 months moving average corresponding to it, i.e., in a multiplicative model we obtain.

$$\frac{\text{Original value}}{\text{Moving Average value}} \times 100 = \frac{TSCI}{TC} = SI \times 100$$

(iii) Arrange these percentage season-wise for all the years. Average these percentages. These values are the seasonal indices which are almost free from irregular variations due to averaging.

(iv) Add these indices. If the sum is not 1200 or 400 for monthly or quarterly figures multiply each value by the correction factor (c.f.).

$$\text{c.f.} = \frac{1200}{\text{Sum of monthly indices}} \text{ or } \frac{400}{\text{Sum of quarterly indices}}$$

This gives ratio to moving average seasonal indices.

Problems

5. Calculate seasonal indices by the ratio-to-moving average method from the following data :

Barley Prices in Rupees Per Quintal				
	Quarter			
Year	**I**	**II**	**III**	**IV**
2004	68	62	63	78
2005	75	58	56	72
2006	60	63	67	93
2007	54	59	56	90
2008	59	55	58	65

Sol :

Year	Quarters	Barley Prices (Rs.)	4-Figure Moving Total	2-Figure Moving Total	4-Figure Moving	Given Figures as % of Moving Average $\frac{\text{col. 3}}{\text{col. 6}}$
1	2	3	4	5	6	7
2004	Q_I	68	–	–	–	–
	Q_{II}	62	–	–	–	–
			271	–	–	–
	Q_{III}	63		549	68.63	(63/68.63) × 100 = 91.79
			278			–
	Q_{IV}	78		552	69.00	(78/69) × 100 = 113.04
			274			–
2005	Q_I	75		541	67.63	(75/67.63) × 100 = 110.90
			267			
	Q_{II}	58		528	66.00	(58/66) × 100 = 87.88
			261			
	Q_{III}	56		507	63.38	(56/63.38) ×100 = 88.36
			246			
	Q_{IV}	72		497	62.13	(72/62.13) × 100 = 115.89
			251			
2005	Q_I	60		513	64.13	(60/64.13) × 100 = 93.56
			262			
	Q_{II}	63		545	68.13	(63/68.13) × 100 = 92.47
			283			
	Q_{III}	67		560	70.00	(67/70) × 100 = 95.71
			277			
	Q_{IV}	93		550	68.75	(93/68.75) × 100 = 135.27

			273			
2006	Q_I	54		535	66.88	(54/66.88) × 100 = 80.74
			262			
	Q_{II}	59		521	65.13	(59/65. 13) × 100 = 90.59
			259			
	Q_{III}	56		523	65.38	(56/65.38) × 100 = 85.65
			264			
	Q_{IV}	90	260	524	65.50	(90/65 .50) × 100 = 137.50
			260			
2007	Q_I	59		522	65.25	(59/65.25) × 100 = 90.423
			262			
	Q_{II}	55		499	62.38	(55/62.38) × 100 = 88.17
			237			
	Q_{III}	38	–			
	Q_{IV}	65	–			

Calculation of Season Index (Trend Eliminated Values)

	Quarter			
Year	**I**	**II**	**III**	**IV**
2004	–	–	91.79	113.04
2005	110.90	87.88	88.36	115.89
2006	93.56	92.47	95.71	135.27
2007	80.74	90.59	85.65	137.50
2008	90.42	88.17	–	–
Total	375.62	359.11	361.51	501.17
Seasonal Average	93.9	89.77	90.375	125.3

Total of quarterly averages = 399.345

Adjusted season indices $= \frac{93.9 \times 400}{399.345} = 94, \quad \frac{87.77 \times 400}{399.345}$

$= 89.91,$

$\frac{90.375 \times 400}{399.345} = 900.52, \quad \frac{125.3 \times 400}{399.345}$

$= 125.50$

6. Calculate seasonal indices by the ratio to moving average method, from that following data :

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2006	**68**	**62**	**61**	**63**
2007	**65**	**58**	**66**	**61**
2008	**68**	**63**	**63**	**67**

Sol :

Calculate of Seasonal Indices by Ratio to Moving Average' Method

Year	Quarter	Given figures	4-figure moving totals	2-figure moving totals	4-figure moving average	Given figure as % of moving average
2006	I	68				
	II	62				
			254			
	III	61		505	63.186	96.63
			251			
	IV	63		498	62.260	101.19
			247			
2007	I	65		499	62.375	104.21
			252			
	II	58		502	62.750	92.43
			250			
	III	66		503	62.875	104.97
			253			
	IV	61		511	63.875	95.50
			258			
2008	I	68		513	64.125	106.04
			255			
	II	63		516	64.500	97.67
			261			
	III	63				
	IV	67				

Calculation of Seasonal Index

Percentage to Moving Average

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
2006	—	—	96.63	101.20
2007	104.21	92.43	104.97	95.50
2008	106.04	97.56	—	—
Total	210.25	190.10	201.60	196.70
Average	105.125	95.05	100.80	98.35
Seasonal Index	105.30	95.21	100.97	98.52

Arithmetic average of averages $= \frac{399.32}{4} = 99.83$

By expressing each quarterly average as percentage of 99.83, we will obtain seasonal in-dices.

Seasonal index of 1 st Quarter $= \frac{105.125}{99.83} \times 100 = 105.30$

$$\text{Seasonal index of 2nd Quarter} = \frac{95.05}{99.83} \times 100$$

$$= 95.21$$

$$\text{Seasonal index of 3rd Quarter} = \frac{100.80}{99.83} \times 100$$

$$= 100.97$$

$$\text{Seasonal index of 4th Quarter} = \frac{98.35}{99.83} \times 100$$

$$= 98.52$$

Q8. Explain the concept of link relative method.

Ans :

This method is also known as the Pearson's method.

Steps involved

(i) Find out the link relatives of seasonal figures.

$$\text{LR} = \frac{\text{Current season's figure}}{\text{Previous season's figure}} \times 100$$

(ii) Calculate the average of link relative for each season.

(iii) Convert LR into chain relatives (CR) on the basis of first season, taking the chain relative of first season as 100.

$$\text{CR} = \frac{\text{LR of the season} \times \text{CR of the previous season}}{100}$$

For example if the data are given on monthly basis the CR for January is 100 and CR for February

$$= \frac{\text{LR of February} \times \text{CR of the previous season, i.e., January}}{100}$$

iv) Get the new chain relative for January on the basis of December CR which would be

$$\text{New CR (January)} = \frac{\text{LR of January} \times \text{CR of December}}{100}$$

v) Make adjustment by subtracting a correction factor from each chain relative.

$$\text{Correction factor (k)} = \frac{\text{New CR for January (or 1st Quarter)} - \text{Old change relative, i.e., 100}}{12 \text{(or 4)}}$$

vi) Express the corrected chain relatives to total 1200 or (400) by expressing them as percentages of their averages, by multiplying each of them by a constant factor.

$$K = \frac{1200}{\text{Sum of the corrected monthly chain relatives}}$$

$$= \frac{400}{\text{Sum of the corrected monthly chain relatives}}$$

PROBLEMS

7. Calculate seasonal indices by the method of link relatives for the following data :

Quarterly figures for Five Years

Year	2004	2005	2006	2007	2008
I	45	48	49	52	60
II	54	56	63	65	70
III	72	63	70	75	83
IV	60	56	65	72	86

Sol :

Calculation of link relatives of the seasonal figures are given in the table below :

Year	Quarter			
	I	II	III	IV
2004	–	120	133.33	83.33
2005	80.00	116.67	112.50	88.89
2006	87.00	128.57	111.11	92.86
2007	80.00	125.00	115.38	96.00
2008	83.33	116.67	118.57	103.63
Mean	82.708	121.382	118.78	92.938
Chain Relatives	100.00	$\frac{100 \times 121.382}{100}$	$\frac{121.382 \times 118.178}{100}$	$\frac{143.7 \times 93}{100}$
Total		= 121.382	= 143.447	= 133.317
Adjusted Chain Realties	100.00	118.816	138.315	125.619 = 482.80
Seasonal indices	= 82.86	= 98.45	= 114.61	= 104.09

Chain relative for first quarter calculated on the basis of chain relative of the last quarter is $\frac{133.317 \times 82.708}{100}$ = 110.264. Adjustment factor for chain relative for 2nd, 3rd and 4th quarter are d, 2d and 3d respectively where d = $\frac{110.264 - 100}{4}$ = 2.566.

8. Apply the method of link relatives to the following data and calculate seasonal indices.

Quarter	Quarterly Figures				
	1995	**1996**	**1997**	**1998**	**1999**
I	6.0	5.4	6.8	7.2	6.6
II	6.5	7.9	6.5	5.8	7.3
III	7.8	8.4	9.3	7.5	8.0
IV	8.7	7.3	6.4	8.5	7.1

Sol :

Calculation of Seasonal Indices by the Method of Link Relatives

Year	Quarter			
	I	**II**	**III**	**IV**
1995	–	108.3	120.0	111.5
1996	62.1	146.3	106.3	86.9
1997	93.2	95.6	143.1	68.8
1998	112.5	80.6	129.3	113.3
1999	77.6	110.6	109.6	88.8

Arithmetic average	$\frac{345.4}{4} = 86.35$	$\frac{541.4}{5} = 108.28$	$\frac{608.3}{5} = 121.66$	$\frac{469.3}{5} = 93.86$
Chain relatives	100	$\frac{100 \times 108.28}{100}$ = 108.28	$\frac{121.66 \times 108.28}{100}$ = 131.73	$\frac{93.86 \times 131.73}{100}$ = 123.65
Corrected chain relatives	100	108 – 1.675 = 106.605	131.73 – 3.35 = 128.38	123.64 – 5.025 = 118.615
Seasonal indices	$\frac{100 \times 100}{113.4}$ = 88.18	$\frac{106.605}{113.4} \times 100$ = 94.01	$\frac{128.38}{113.4} \times 100$ = 113.21	$\frac{118.615}{113.4} \times 100$ = 104.60

The calculations in the above table are explained below :

Chain relative of the first quarter (on the basis of first quarter) = 100

Chain relative of the first quarter (on the basis of the last quarter) = $\frac{86.35 \times 123.64}{100} = 106.7$

The difference between these chain relatives = 106.7 – 100 = 6.7.

Difference per quarter = $\frac{6.7}{4}$ = 1.675.

Adjusted chain relatives are obtained by subtracting 1 × 1.675, 2 × 1.675, 3 × 1.675 from the chain relatives of the 2nd, 3rd and 4th quarters respectively.

Average of corrected chain relatives

$$\frac{100+106.605+128.38+118.615}{4} = \frac{453.6}{4} = 113.4$$

$$\text{Seasonal variation index} = \frac{\text{Correct chain relatives}}{113.4} \times 100$$

3.3 Methods

3.3.1 Semi Averages

Q9. What are the various methods of time series? Explain in detail semi average method. With an example.

Ans : **(Imp.)**

There are four methods for determining trend in time series :

i) Semi-Average Method,

ii) Freehand (or Graphical) Method,

iii) Moving Average Method.

iv) Least Squares Method;

i) Semi-Average Method

The following procedure is followed for semi-average method,

- The entire time series is classified into two equal parts with respect to time. For even period, equal split. For odd period, equal parts obtained by omitting middle period.
- Compute the arithmetic mean of time series values for each half separately. These means are called semi-averages.
- Semi averages are plotted as points against the middle point of the respective time period covered by each part.
- The line joining these points gives the straight line trend fitting the given data.

Merits of Semi-Average Method

The following are some of the merits of semi-average method,

- Objectivity
- Ease of apply and understandability
- Extend both ways the line i.e., we can get past and future estimates.

Demerits of Semi-Average Method

Some of the demerits of semi-average method are,

- Linear trend assumption may not exist.
- A men may be questioned.
- Thus, values of trend are not precise and reliable.

Example

Using the following data, fit a trend line by using the method of semi-averages,

Year	1996	1997	1998	1999	2000	2001	2002
Output	700	900	1100	900	1300	1000	1600

Sol :

Step 1

The data provided in the problem is of seven years i.e., (an odd number). Thus, the middle year [1999] shall be ignored and the remaining years are divided into two equal time periods and their arithmetic averages is computed as follows,

Average of the first three years $= \frac{700+900+1100}{3} = \frac{2700}{3} = 900$

Average of the last three years $= \frac{1300+1000+1600}{3} = \frac{3900}{3} = 1300$

Therefore, the semi-averages are 900 and 1300

Step 2

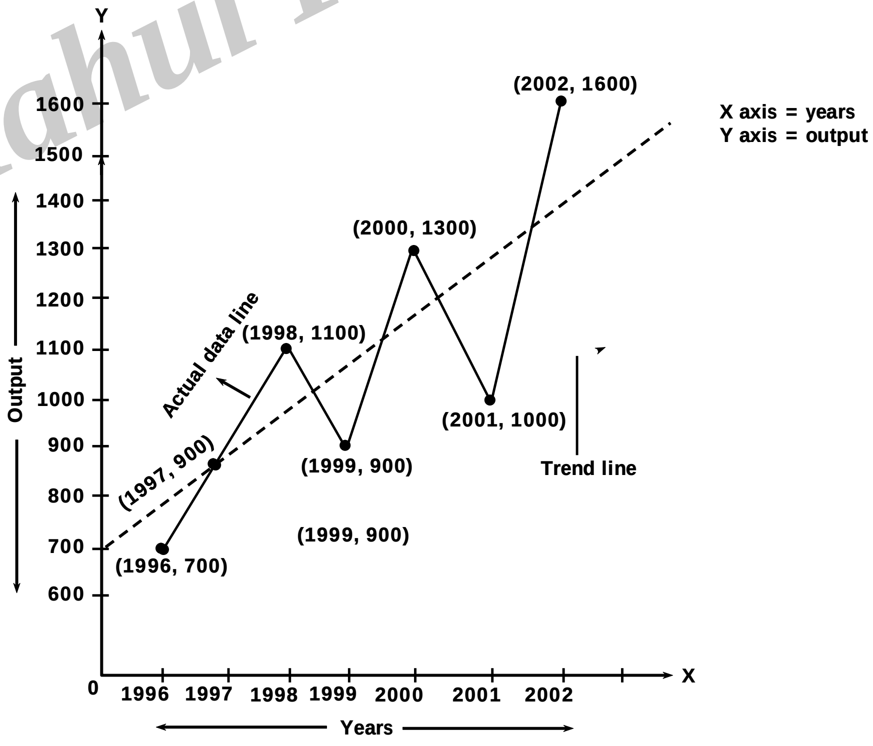
The next step is to plot the semi-averages against the mid-point (middle year) of each time period. Thus, it would , be year 1997 and 2001 respectively.

Step 3

The plotted points are joined in order to derive the trend line using the semi average method.

Step 4

The original data and the trend line is plotted on a graph as follows,



3.3.2 Free Hand Curve

Q10. What is Free Hand Curve? Explain with illustration.

Ans :

A trend is determined by just inspecting the plotted points on a graph sheet. Observe the up and down movements of the-points. Smooth out the irregularities by drawing a freehand curve or line through the scatter points. The curve so drawn would give a general notion of the direction of the change. Such a freehand smoothed curve eliminates the short- time swings and shows the long period general tendency of the changes in the data.

Drawing a smooth freehand curve requires a personal skill and judgement. The drawn curve should pass through the plotted points in such a manner that the variations in one direction are approximately equal to the variation in other direction. Different persons, however, drawn different curves at different directions, with different slopes and in different styles. This may lead to different conclusions. To overcome these limitations, we can use the semi-average method of measuring the trend.

Merits

(i) It is very simple.

(ii) It does not involve any calculations.

(iii) It is very flexible and can be used irrespective of whether the trend is linear or curvi-linear.

(iv) If used by experienced statisticians, it is a better tool to study trend movement compared to other methods using rigid mathematical formulae.

Limitations

(i) It is very subjective. Different persons may draw different lines and reach different conclusions from the same data. Hence, it is not a good forecasting tool.

(ii) If properly attempted, it is very time-consuming effort.

(iii) It requires high levels of experience and expertise to effectively use this method.

Example

Fit a trend line to the following data by the freehand method,

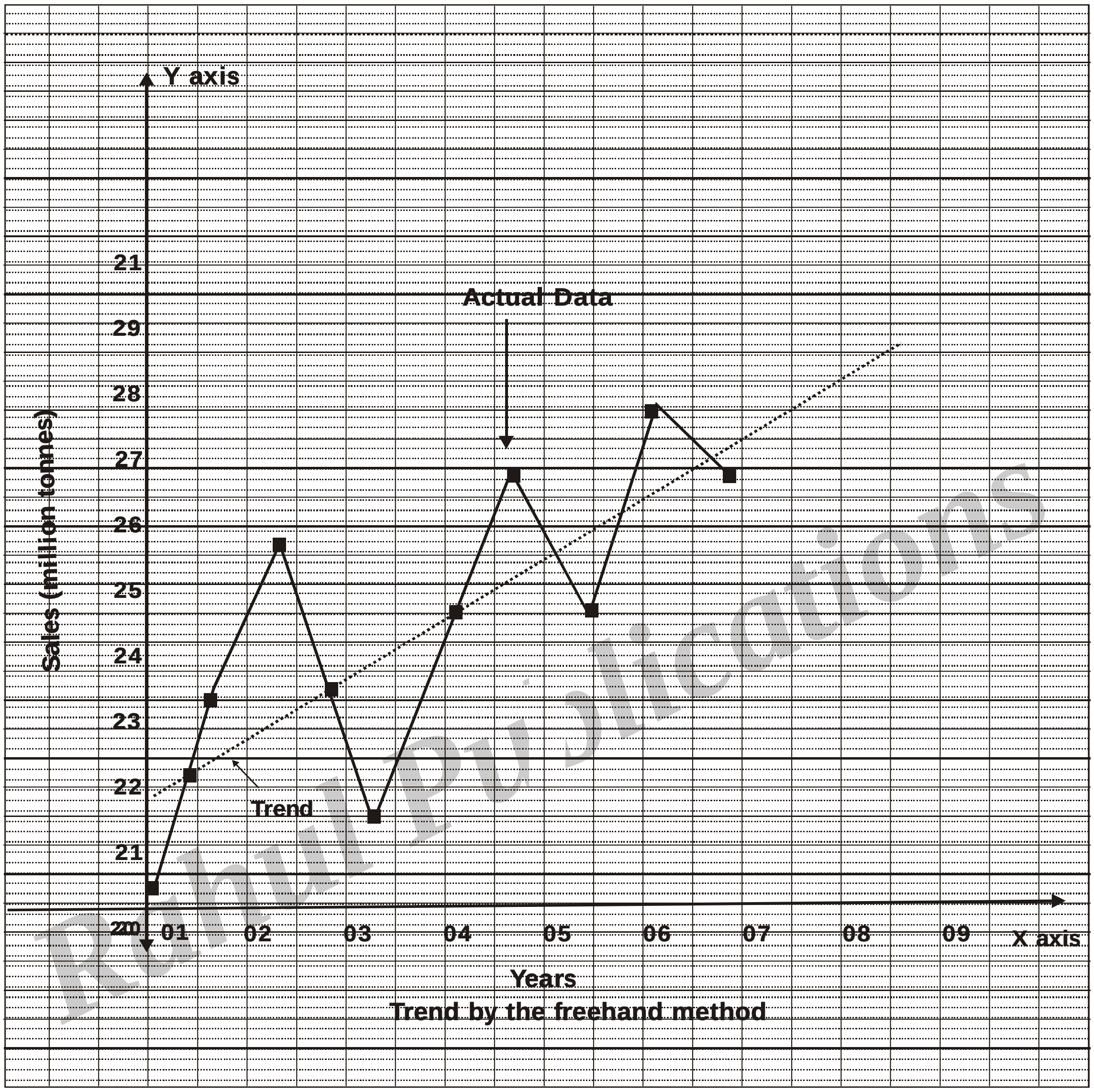
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sales (milllion tonnes)	19	22	24	20	23	25	23	26	25

Sol :

Steps

1. Time series data is plotted on the graph
2. The direction of the trend is examined on the basis of the plotted data (dots)
3. A straight line is drawn which shows the direction of the trend.

The actual data and the trend line are shown in the following graph.



3.3.3 Moving Averages

Q11. Discuss the method of moving averages in measuring trend. What are it merits and limitations of moving average method?

Ans : **(Imp.)**

In moving average method, the average value for a number of years (month or weeks) is secured and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average.

The effect of averaging is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend.

The period of moving average is decided in the light of the length of the cycle. More applicable to data with cyclical movements.

Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}...$$

Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}....$$

Methods

The following two methods are followed in moving averages,

a) Odd Yearly Method

i) Calculate 3/5...yearly totals

ii) Now compute 3/5 yearly average by dividing the totals calculated in step (i) by the respective number of years, i 3/5/...

iii) Short term oscillations are calculated using the formula, $Y - Y_C$

Where, Y - Actual value and Y - Estimated value.

b) Even Yearly Method

Example : 4 years

i) Calculate 4 yearly moving totals and place at the centre of middle two years of the four years considered.

ii) Divide 4 yearly moving totals by 4 to get 4 yearly average.

iii) Take a 2 period moving average of the moving average which gives the 4 yearly moving average centered.

Merits

The merits of moving average are as follows,

a) Of all the mathematical methods of fitting a trend, this method is the simplest.«

b) The method is flexible so that even if a few more observations are to be added, the entire calculations are not changed.

c) If the period of the moving average happens to coincide with the period of the cycle, the cyclical fluctuations are automatically eliminated.

d) The shape of the curve in case of moving average method is determined by the data rather than the statisticians choice of mathematical function.

Limitations

The following are the limitations of moving averages,

a) Trend values cannot be computed for all the years. For example, in a 5 yearly moving we cannot compute trend values for the first two and the last two years.

b) It is difficult to decide the period of moving average since there is no hard and fast rule for the purpose.

c) Moving average cannot be used in forecasting as it is not represented by any mathematical function.

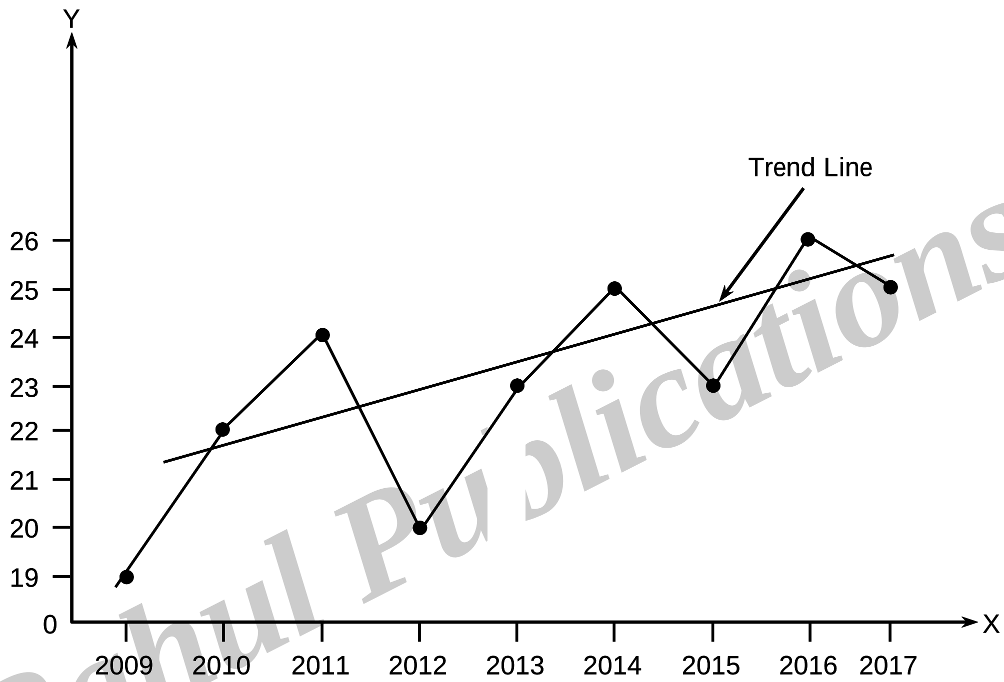
d) When the trend is not linear, the moving average lies either above or below the true sweep of the data.

PROBLEMS

9. Fit a trend line to the following data by the freehand method.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Sales (Rs.)	19	22	24	20	23	25	23	26	25

Sol : **(June-19)**



10. From the following data fit a trend line by the method of Semi-Average.

Year :	2012	2013	2014	2015	2016	2017
Output :	20	16	24	30	28	32

Sol : **(Jan.-21, Imp.)**

The trend value of first 3 years calculated as follows

$$\frac{20+16+24}{3} = \frac{60}{3} = 20.$$

The trend value of last 3 years are calculated as follows

$$\frac{30+28+32}{3} = \frac{90}{3} = 30$$

Therefore, the Semi Average are 20 and 30.

11. The following table gives the annual sales (in Rs.'000) of a commodity :

Year	Sales
1990	710
1991	705
1992	680
1993	687
1994	757
1995	629
1996	644
1997	783
1998	781
1999	805
2000	805

Determine the trend by calculating the 5-yearly moving average.

Sol :

Calculation of Trend by 5-Yearly Moving Average

Year	Sales	5-Yearly Moving Total	5-Yearly Moving Average
1990	710		
1991	705		
1992	680	710 + 705 + 680 + 687 + 757 = 3539	$\frac{3539}{5} = 707.8$
1993	687	705 + 680 + 687 + 757 + 629 = 3458	$\frac{3458}{5} = 691.6$
1994	757	680 + 687 + 757 + 629 + 644 = 3397	$\frac{3397}{5} = 679.4$
1995	629	687 + 757 + 629 + 644 + 783 = 3500	$\frac{3500}{5} = 700$
1996	644	757 + 629 + 644 + 783 + 781 = 3594	$\frac{3594}{5} = 718.8$
1997	783	629 + 644 + 783 + 781 + 805 = 3642	$\frac{3642}{5} = 728.4$
1998	781	644 + 783 + 781 + 805 + 872 = 3885	$\frac{3885}{5} = 777$
1999	805		
2000	872		

12. Calculate three year moving average for the following data:

Year :	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
Value :	242	250	252	249	253	255	251	257	260	265	262

Sol :

Calculation of Trend by 3 years moving average

Year	Value	3 year moving total	3 year moving average
1950	242	–	–
1951	250	242 + 250 + 252 = 744	744/3 = 248.00
1952	252	250 + 252 + 249 = 751	751/3 = 250.33
1953	249	252 + 249 + 253 = 754	754/3 = 251.33
1954	253	249 + 253 + 255 = 757	757/3 = 252.33
1955	255	253 + 255 + 251 = 759	759/3 = 253.00
1956	251	255 + 251 + 257 = 763	763/3 = 254.33
1957	257	251 + 257 + 260 = 768	768/3 = 256.00
1958	260	257 + 260 + 265 =782	782/3 = 260.67
1959	265	260 + 265 + 262 = 787	787/3 = 262.33
1960	262	–	–

13. Calculate the trend values by the method of moving average, assuming a four yearly cycle, from the following data relating to sugar production in India.

Year	Sugar Production (Lakh Tonnes)	Year	Sugar Production (Lakh Tonnes)
1971	37.4	1977	48.4
1972	31.1	1978	64.6
1973	38.7	1979	58.4
1974	39.5	1980	38.6
1975	47.9	1981	51.4
1976	42.6	1982	84.4

Sol :

Calculation of 4 years moving average

Year	Sugar Prod (lakh Tones)	4 yearly Moving Total	4 yearly moving Average	2 period Moving Total Average	Centered Moving
1971	374	-	-	-	-
1972	31.1	-	-	-	-
		37.4 + 31.1 + 38,7 + 39.5 – 146.7	36.675		
1973	38.7			36.675 + 39.300 = 75.975	37.99
		31.1 + 38.7 + 39.5 + 47.9 = 157.2	39.300		
1974	39.5			39.300 + 42.175 = 81.475	40.75
		38.7 + 39.5 + 47.9 + 42.6 = 42.175	42.175		
1976	47.9			42.175 + 50.875 = 93.05	46.525
		47.9 + 42,6 + 48.4 + 64.6 = 203.5	50.875		
1977	48.4			50.875 + 53.500 = 104.375	52.19
		42.6 + 48.4 + 6 + 58.4 = 214.0	53:500		
1978	64.6			53.500 + 52.500 = 106.000	53.00
		48.4 + 64.6 + 58.4 + 38.6 = 210.00	52.500		
1979	58.4			52.500 + 53.250 = 105.750	52.88
		64.6 + 58.4 + 38.6 = 210.0	52.500		
1980	38.6			53.500 + 58.200 = 111.450	55.73
1981	51.4	-	-	-	-
1982	84.4	-	-	-	-

14. Find the 4 yearly moving averages from the following data.

Year	2008	209	2010	2011	2012	2013	2014	2015	2016	2017
Production (in Tonnes)	150	170	196	180	190	216	248	280	300	320

Sol : **(June-18, Imp.)**

Calculation of 4 Years Moving Average

Year	Production	4-Yearly Moving Total	4-Yearly Moving Avg.	2 Points Moving total	2 Years Moving Avg.
2008	150				
2009	170	(150 + 170 + 196 + 180)	$\frac{696}{4} = 174$		
	$\longrightarrow$	696	$\longrightarrow$	358	179
2010	196	(170 + 196 + 180 + 216)	184		
	$\longrightarrow$	736	$\longrightarrow$	379.5	189.75
2011	180		195.5		
	$\longrightarrow$	782	$\longrightarrow$	404	202
2012	190		208.5		
	$\longrightarrow$	834	$\longrightarrow$	442	221
2013	216		233.5		
	$\longrightarrow$	934	$\longrightarrow$	494.5	247.25
2014	248		261		
	$\longrightarrow$	1044	$\longrightarrow$	548	274
2015	280	(248 + 280 + 300 + 320)			
	$\longrightarrow$	1148	287		
2016	300				
2017	320				

15. From the following data, calculate trend values using Four Yearly Moving Averages.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	506	620	1036	673	588	696	1116	738	663

Sol : **(June-19)**

Calculation of Trend by 4 Years Moving Average

Years	Production	4 Years Total	4 Years Average	2 Point Moving Total	4 Yearly Moving Average Centered
2009	506				
2010	620				
		2,835	708.75		
2011	1036			1438	719
		2,917	729.25		
2012	673			1477.5	738.75
		2,993	748.25		
2013	588			1516.5	758.25
		3073	768.25		
2014	696			1552.75	776.375
		3138	784.5		
2015	1116			1587.75	793.875
		3213	803.25		
2016	738				
2017	663				

16. Calculate the 3 yearly the 5 yearly moving averages for the following time series.

Year :	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Production (in quintals)	500	540	550	530	520	560	600	640	620	610	640

Sol : **(Jan.-21, Imp.)**

Calculation of 3 Years and 5 Years Moving Averages

Years	Production	3 year Total	3 years Average	5 years Total	5 years Average
2006	500	–	–		
2007	540	1590	$\frac{1590}{3} = 530$	–	
2008	550	1620	540	2640	$\frac{2640}{5} = 528$
2009	530	1600	533.33	2700	540
2010	520	1610	536.66	2760	552
2011	560	1680	560	2850	570
2012	600	1800	600	2940	588
2013	640	1860	620	3030	606
2014	620	1870	623.33	3110	622
2015	610	1870	623.33		
2016	640	–			

3.3.4 Least Square Method

Q12. Define least square method. Explain merits and demerits of least square method.

(OR)

What is least square method and explain its advantages and disadvantages?

Ans : **(Imp.)**

Least square method is the most widely used method and provides us with a mathematical device to obtain an objective fit to the trend of a given time series. This method is so called because a trend line computed by this method is such that the sum of the squares of the deviation between the original data and the corresponding computed trend values is minimum. This method can be used to fit either a straight line trend or a parabolic trend.

The straight line trend equation is in the form of $Y = a + bX$

Where, Y denotes the trend value of the dependent variable

X denotes the independent variable.

a and b are constants.

The values of a and b are obtained by solving the following normal equations.

$$\Sigma Y = Na + b\Sigma X$$

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

Where, N represents the number of years in the series.

When $\Sigma X = 0$ the above normal equations are simplified to

$$a = \frac{\sum Y}{N}$$

$$b = \frac{\sum XY}{\sum X^2}$$

By substituting a and b values in straight line trend equation $Y = a + bX$, we get the straight line equation which can be used for estimation of future values.

Merits of Least Squares Method

The following are the merits of least squares method,

1. The method of least squares is a mathematical method of measuring trend and is free from subjectiveness.
2. This method provides the line of best fit since it is this line from where the sum of positive and negative deviations is zero and the sum of square of deviations is the least.
3. This method enables us to compute the trend values for all the given time periods in the series.
4. The trend equation can be used to estimate the values of the variable for any given time period 'f' in future and the forecasted values are quite reliable.
5. This method is the only technique which enables us to obtain the rate of growth per annum for yearly data in case of linear trend.

Demerits of Least Squares Method

Some of the demerits of least squares are as follows,

1. Fresh calculations become necessary even if a single new observation is added.
2. Calculations required in this method are quite tedious and time consuming as compared with other methods.
3. Future predictions based on this method completely ignore the cyclical, seasonal and erratic fluctuations.
4. This method cannot be used to fit growth curves, gomper t_z curve, logistic curve etc. to which most of the business and economic time series conform.

PROBLEMS

17. Obtain the straight line trend equation for the following data by the method of the least square. Tabulate the trend values.

Year :	2010	2011	2012	2013	2014	2015	2016
Sale (in '000 units)	140	144	160	152	168	176	180

Sol : **(Jan.-21)**

$Y_C = a + bx$

$a = \frac{\Sigma y}{n} \quad b = \frac{\Sigma xy}{\Sigma x^2}$

Fitting of straight line trend

Year	Sales	x	xy	x^2	$y_c = a + bx$
2010	140	– 3	– 420	9	139.42
2011	144	– 2	– 288	4	146.28
2012	160	– 1	– 160	1	153.14
2013	152	0	0	0	160
2014	168	1	168	1	166.86
2015	176	2	352	4	173.72
2016	180	3	540	9	180.58
	1120		192	28	

$a = \frac{1120}{7} = 160$

$b = \frac{192}{28} = 6.86$

$y_c = 160 + 6.86(x)$

For 2010 $= 160 + 6.86(-3)$

$160 - 20.58 = 139.42$

For 2011 $= 160 + 6.86(-2)$

$160 - 13.72 = 146.28$

For 2012 $= 160 + 6.86(-1)$

$160 - 6.86 = 153.14$

For 2013 $= 160 + 6.86(0)$

$160 - 0 = 160$

For 2014 $= 160 + 6.86(1)$

160 + 6.86 = 166.86

For 2015 = 160 + 6.86(2)

160 + 13.72 = 173.72

For 2016 = 160 + 6.86(3)

= 160 + 20.58

= 180.58

18. Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

Sol : **(June-19, Imp.)**

Fitting of Straight line trend

Years	Y	Years – Middle Year = X Years – 2014 = X	X^2	XY	Y_c
2011	77	– 3	9	– 231	83
2012	88	– 2	4	– 176	85
2013	94	– 1	1	– 94	87
2014	85	0	0	0	89
2015	91	+ 1	1	91	91
2016	98	+ 2	4	196	93
2017	90	+ 3	9	270	95
	623	0	28	56	623

$y_c = a + bx$

$a = \frac{\Sigma y}{N} = \frac{623}{7} = 89$

$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{56}{28} = 2$

Calculation of Trend Values (Y_C)

2011 $Y_c = a + bx = 89 + [2][-3] = 89 - 6 = 83$

2012 $Y_c = a + bx = 89 + [2][-2] = 89 - 4 = 85$

2013 $Y_c = a + bx = 89 + [2][-1] = 89 - 2 = 87$

2014 $Y_c = a + bx = 89 + [2][0] = 89 + 0 = 89$

2015 $Y_c = a + bx = 89 + [2][1] = 89 + 2 = 91$

2016 $Y_c = a + bx = 89 + [2][2] = 89 + 4 = 93$

2017 $Y_c = a + bx = 89 + [2][3] = 89 + 6 = 95$

19. Production figure of a Textile Industry are as follows.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in '000 units)	12	10	14	11	13	15	16

For the above data

(i) Determine the straight line equation under the Least Square Method.

(ii) Find the Trend Values and show the trend line on a graph paper.

Sol : **(June-18)**

Fitting of straight line equation

Year	Production (y)	x	x^2	xy	Trend Values $y_c = a + bx$
2011	12	–3	9	–36	10.75
2012	10	–2	4	–20	11.5
2013	14	–1	1	–14	12.25
2014	11	0	0	0	13
2015	13	1	1	13	13.75
2016	15	2	4	30	14.5
2017	16	3	9	48	15.25
N = 7	91	0	28	21	

We know, Trend value,

$$y_c = a + bx$$

$$a = \frac{\Sigma y}{n} = \frac{91}{7} = 13$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{21}{28} = 0.78$$

For 2011

13 + 0.75 (–3)

= 13 – 2.25

= 10.75.

For 2012

13 + 0.75 (–2)

= 13 – 1.5

= 11.5

For 2013

$13 + 0.75\,(-1)$

$= 13 - 0.75$

$= 12.25$

For 2014

$13 + 0.75\,(0)$

$= 13 + 0$

$= 13$

For 2015

$13 + 0.75\,(1)$

$= 13 + 0.75$

$= 13.75$

For 2016

$13 + 0.75\,(2)$

$= 13 + 1.5$

$= 14.5$

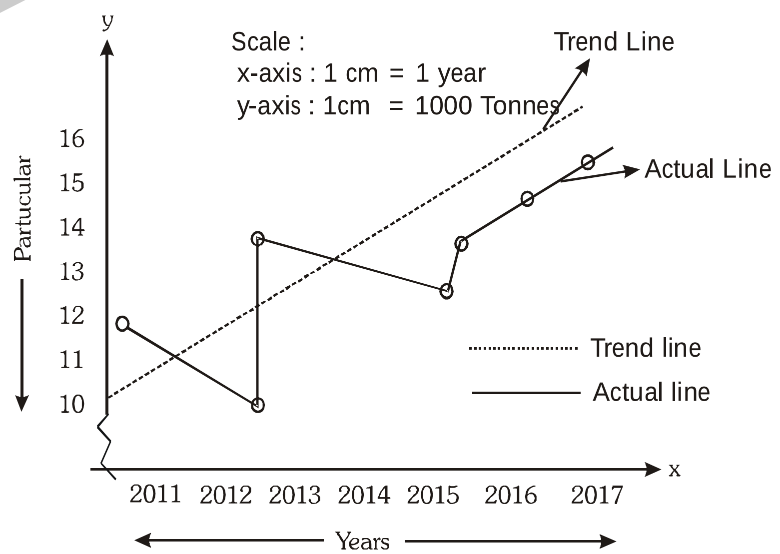
For 2017

$13 + 0.75\,(3)$

$= 13 + 2.25$

$= 15.25$

ii) Find the Trend values and show the trend Line on a graph Paper.



20. Given below is the data of production of a certain company in lakhs of units.

Year	1995	1996	1997	1998	1999	2000	2001
Production	15	14	18	20	17	24	27

Compute the Linear trend by the method of least squares. Predict the production for the year "2004".

Sol :

The Normal Equation are,

$$\Sigma Y = Na + b\Sigma X$$

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

By solving these equations, we get a and b. This method is called the method of least squares.

Fitting of straight line trend

Year	Production Y	from 1998 X	X^2	XY
1995	15	– 3	9	– 45
1996	14	– 2	4	– 28
1997	18	– 1	1	– 18
1998	20	0	0	0
1999	17	1	1	17
2000	24	2	4	48
2001	27	3	9	81
Total	**135**	**0**	**28**	**55**

$$y_c = a + bx$$

Since $\Sigma X = 0$, the above equations simplify to,

$$a = \frac{\Sigma Y}{N} = \frac{135}{7} = 19.29$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{55}{28} = 1.96$$

$\therefore$ The straight line trend, $Y_e = 19.29 + 1.96X$.

Prediction of Production for the Year 2004

For year 2004,

$X = 6$ (deviation from mid year 1998).

$\therefore \quad Y_c = 19.29 + 1.96\ (6)$

$= 19.29 + 11.76$

$= 31.05$

The production of the company for the year 2004 is predicated to be 31.05 lakhs of units.

21. The sales of a company in lakhs of rupees for the years 1994-2001 are given below:

Year	1994	1995	1996	1997	1998	1999	2000	2001
Production	550	560	555	585	540	525	545	585

Compute the Linear trend by the method of least squares. Predict the production for the year "2002".

Sol :

The model of least squares helps in solving this problem.

The straight line trend is $y_c = a + bX$

By solving these normal equations, we get a and b.

Normal equation are,

$\Sigma Y = Na + b\Sigma X$

$\Sigma XY = a\Sigma X + b\,\Sigma X^2$

Since there are even number of observations, deviations are taken from both of the middle periods i.e., 1997 and 1998.

Fitting of straight line trend

Year	Sales Y	From mid years X	X^2	XY
1994	550	–7	49	– 3850
1995	560	–5	25	–2800
1996	555	–3	9	–1665
1997	585	–1	1	–585
1998	540	1	1	540
1999	525	3	9	1575
2000	545	5	25	2725
2001	585	7	49	4095
N = 8	**$\Sigma Y = 4445$**	**$\Sigma X = 0$**	**$\Sigma x^2 = 168$**	**$\Sigma XY = 35$**

Since $\Sigma Y = 0$, the normal equation simplify to,

$$a = \frac{\Sigma Y}{N} = \frac{4445}{8} = 555.63$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{35}{168} = 0.21$$

$\therefore$ The straight line trend is $Y_e = 555.63 + 0.21X$

Estimation of Sales for the Year 2002

For year 2002, deviation X = 9,

$\therefore$ $Y = 555.63 + 0.21(9)$

$= 555.63 + 1.89$

$= 557.52$

For year 2002, the sales of the company is estimated to be ` 557.52 lakhs.

3.4 Deseasonalisation of Data

Q13. What do you mean by Deseasonalisation of Data?

Ans :

One of the objectives of calculating seasonal indices is to eliminate the impact of seasonal factors on the given time series values, the process of elimination of effect of seasonal variations from the given data is called 'Depersonalization. 'Depersonalization helps in proper interpretation of data. The process depersonalization is dependent on whether Additive model or Multiplicative model is applicable to the given data.

In an additive model where Y = T + S + C + I, Deseasonalise Values = Y-S = T + C + I. In a multiplicative model where Y = T × S × C × I.

Deseasonalised Values = Y/S = T x C x I.

Problems

22. Deseasonalise the following data by (a) Absolute Seasonal Variation and (b) Deseasonalisation by Seasonal Index:

Month	Jan	Feb	Mar	Apr	May	June
Sales (`)	167/7	83.4	134.7	116.7	77.1	101.4
Seasonal Variation (Adjusted)	29.4	-27.3	25.2	9.3	-33.9	-24.3
Seasonal Index	125	76.2	122,1	108.4	70.6	78.4
Month	Jul	Aug	Sep	Oct	Nov	Dec
Sales	108.9	113.1	91.5	95.4	148.8	178.5
Seasonal Variation (Adjusted)	-13.2	-15.9	-29.1	-11.1	21.6	70.8
Seasonal Index	86.8	85.3	73.6	90.5	119.1	164

Sol :

Period	Sales (Rs.Cr.)	Seasonal Variation (Adj)	Deseasonalised Sales (Additive Model)	Seasonal Index (Adj)	Deseasonalised Sales (Multiplicative Model)
Jan	167.7	29.4	167.7-29.4 = 1383	125	167.7 × 100/125 = 134.2
Feb	83.4	-27.3	83.4 (-27.3)= 1107	76.2	109.4
Mar	134.7	25.2	109.5	122.1	110.3
Apr	116.7	9.3	107A	108.4	107.7
May	77.1	-33.9	111.0	70.6	109.2
Jun	101.4	-24.3	125.7	78.4	129.3
Jul	108.9	-13.2	122.1	86.8	125.5
Aug	113.1	-15.9	129.0	85,3	132.6
Sep	91.5	-29.1	120.6	73.6	124.3
Oct	95.4	-11.1	106.5	90.5	105.4
Nov	148.8	21.6	127.2	119.1	124,9
Dec	178.5	70.8	107.7	164	108.8

23. The seasonal indices of the sales of garments of a particular type in a certain shop are given below:

Period	Quarter I	Quarter II	Quarter III	Quarter IV
Seasonal Indices	**97**	**85**	**83**	**135**

If the total sales in the first quarter of a year be worth Rs. 15,000 and sales are expected to rise by 4% in each quarter, determine how much worth of garments of this type be kept in stock by the shop owner to meet the demand for each of the three quarters of the year.

Sol :

Normal Expected Sales in any quarter = 124% of sales in previous quarter. However, stock to be kept in any quarter has to adjust for seasonal variations. Thus, stock to be maintained in any quarter = Normal Expected Sales × Seasonal Index.

Thus, value of stock =

Quarter	Normal Expected Sales (Rs)	Seasonal Index	Stock Value (Rs)
Q-I	15,000	97	15,000 x 97% = 14,550
Q-II	15,000 x 104% =15,600	85	15,600 x 85% = 13,260
Q-III	15,600 x 104% = 16,224	83	16,224 x 83% = 13,466
Q-IV	16,224 x 104% = 16,873	135	16,873 x 135% = 22,779

3.5 Uses and Limitations of Time Series

Q14. Explain the uses and limitations of time series.

(OR)

Explain the utility of time series.

(OR)

What are the uses of time series ?

Ans : **(June-19, June-18, Imp.)**

Uses

"Time series analysis has wide spread applications in the areas of business and economics. It is also used in various other disciplines such as natural, social and physical sciences. The utility of time series is as under:

i) It helps in understanding past behavior of the given phenomenon.

ii) It helps in planning future operations and in the formulation of executive and policy decisions.

iii) It helps in evaluation of current achievement by comparing actual achievement with expected achievement.

iv) It helps in forecasting the behavior of the phenomenon in future.

v) It helps compare change in values of different phenomenon at different times.

vi) It helps in isolation of the impact of various factors affecting the time series.

Limitations

a) Most of the methods proposed require to measurement of the concentrations of all the species. This conditions cannot be met in practice.

b) Most of the techniques involve monitoring the concentration relaxation after a perturbation very close to the equilibrium state. The mechanisms deduced by these methods follow pseudo-first-order kinetics.

c) The determination of the mechanism by time series analysis does not afford unique solutions.

d) There are reaction mechanisms that can be kinetically indistinguishable by time series analysis.

Short Question and Answers

1. What is time sereis?

Ans :

Meaning

A time series is a statistical data that are collected, observed or recorded at regular intervals of time. The term time series applies, for example, to the data recorded periodically showing the total annual sales of retail stores, the total quarterly value of construction contracts awarded, the total amount of unfilled orders in durable goods industries at the end of each month, weekly earnings of workers in an industrial town, hourly temperature in a particular city.

Definitions

Some of the important definitions of time series, given by different experts are as under:

i) **According to Morris Hamburg,** "A time series is a set of statistical observations arranged in chronological order."

ii) **According to Patterson,** "A time series consists of statistical data which are collected, recorded observed over successive increments."

iii) **According to Ya-Lun-Chou,** "A time series may be defined as a collection of magnitudes belonging to different time periods, of some variable or composite of variables, such as production of steel, per capita income, gross national product, price of tobacco, or index of industrial production."

iv) **According to Wessel and Wellet,** "When quantitative data are arranged in the order of their occurrence, the resulting statistical series is called a time series."

v) **According to Spiegel,** "A time series is a set of observations taken at specified times, usually at 'equal intervals'. Mathematically, a time series is defined by the values Y_1 Y_2 ... of a variable Y (temperature, closing price of a share, etc.) at times t_1 , t_2 .... Thus Y is a function of t_1 symbolized by Y= F(t)."

vi) **According to Cecil H. Mayers,** "A time series may be defined as a sequence of repeated measurement of a variables made periodically through time".

2. What are the characteristics of time series ?

Ans :

The essentials of Time Series are

i) It must consist of a set of values that are homogeneous. For example, production data for a year and sales data for the next year will not be a time series.

ii) The values must be with reference to time. In other words, in a time series, we have at least 2 variables, with one variable necessarily being time.

iii) The data must be available for a reasonably long period of time.

iv) The gaps between various time values should as far as possible be equal.

v) The values of the second variable should be related to time. For example, the number of people being hired by the BPO industry can be tracked in relation to time. However, if we are talking about average height of students in a class, this data may not haye a significant relationship with time and may not constitute a time series.

3. What are the objectives of time series ?

Ans :

While it is true that past performance does not necessarily guarantee future results, the quality of forecasts that management can make is strongly related to the information that can be extracted and used from past data. Thus, the objective of time series analysis is to interpret the changes in a given variable with reference to the given situation and attempt to anticipate the future course of events. Analysis of Time Series is done with the following objectives:

i) To evaluate past performance in respect of a particular variable.

ii) To make future forecasts in respect of the particular variable.

iii) To chart short term aad long term strategies of the business in respect of the particular variable.

4. Secular Trend or Long Term Movement

Ans :

Secular Trend is the basic tendency of a series to grow, decline, remain constant or fluctuate, over a long period of time. The concept of trend does not include short term changes but is concerned with steady movement over a long period. The long term trend movement is the result of forces that experience change very gradually and continuously over a long period of time. They operate in an evolutionary manner and do not reflect sudden changes. For example, the number of people travelling by Airways has gradually and continuously gone up. The number of infant deaths per thousand children bom is a steadily declining trend. This gradual and continuous movement of trend can be attributed to various factors such as increased investment in infrastructure, opening up of the economy through economic reforms, advances in technology, change in demographic profile etc.

The long term trend helps us in determining the direction of change. The growth factor can be estimated with a fair degree of accuracy. It provides indication of what is ahead for a given series. Elimination of trend from the original data helps in understanding the other elements that influence data in the short run.

5. Seasonal Variations

Ans :

Seasonal variations involve patterns of change within a year that tend to get repeated year after year. The trend is fluctuating and repetitive as it peaks and bottoms out at about the same time of the year, every year. Seasonal variations can be detected only if the data is recorded in smaller units of time such as weekly, monthly or quarterly. There are no seasonal variations in a time series where only annual figures are available. These are the result of such factors that uniformly rise and fall in magnitude. For example, the sale of umbrellas peaks in the months of June and July every year. This is on account of the onset of monsoons. The prices of agricultural commodities fall at the time of harvest. The passenger traffic increases substantially in the summer vacations. In all these examples, the movement of the trend is for a short period of time (season). The same movements are repeated in the coming years. Hence, it is easy to forecast the future.

Characteristics

Based on the above discussion, we can list the following characteristics of seasonal variations.

a) They repeat themselves periodically in less than one year's time.

b) These are results of factors that uniformly and regularly rise and fall in magnitude.

c) These variations are periodic and regular. They can be predicted without much difficulty.

6. Cyclical Variations

Ans :

The cyclical variations in a series are the recurrent variations whose duration is more than one year. Cyclical variations are regular but not uniformly periodic. It is relatively more difficult to predict the future direction of the series. One complete period that normally lasts from seven to nine years, is termed as a cycle. The most common example of Cyclical Variations is the business cycle. The business cycle passes through four phases of Boom, Recession, Depression and Recovery and it may take more than 10 years to complete the cycle.

Characteristics

The characteristics of a cyclical variation are

(a) There are Oscillating movements above and below the secular tread line.

(b) The fluctuations occur over a period greater than one year.

(c) There is no uniformity in the period of recurrence of movement pattern. One cycle may get completed in 3 years while the next cycle for the same date may takes 8 years.

(d) Cyclical fluctuations are more difficult to measure.

7. Semi-Average Method

Ans :

The following procedure is followed for semi-average method,

- The entire time series is classified into two equal parts with respect to time. For even period, equal split. For odd period, equal parts obtained by omitting middle period.
- Compute the arithmetic mean of time series values for each half separately. These means are called semi-averages.
- Semi averages are plotted as points against the middle point of the respective time period covered by each part.
- The line joining these points gives the straight line trend fitting the given data.

Merits of Semi-Average Method

The following are some of the merits of semi-average method,

- Objectivity
- Ease of apply and understandability
- Extend both ways the line i.e., we can get past and future estimates.

Demerits of Semi-Average Method

Some of the demerits of semi-average method are,

- Linear trend assumption may not exist.
- A men may be questioned.
- Thus, values of trend are not precise and reliable.

8. What is Free Hand Curve?

Ans :

A trend is determined by just inspecting the plotted points on a graph sheet. Observe the up and down movements of the-points. Smooth out the irregularities by drawing a freehand curve or line through the scatter points. The curve so drawn would give a general notion of the direction of the change. Such a freehand smoothed curve eliminates the short- time swings and shows the long period general tendency of the changes in the data.

Drawing a smooth freehand curve requires a personal skill and judgement. The drawn curve should pass through the plotted points in such a manner that the variations in one direction are approximately equal to the variation in other direction. Different persons, however, drawn different curves at different directions, with different slopes and in different styles. This may lead to different conclusions. To overcome these limitations, we can use the semi-average method of measuring the trend.

9. Moving average method

Ans :

In moving average method, the average value for a number of years (month or weeks) is secured and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average.

The effect of averaging is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend.

The period of moving average is decided in the light of the length of the cycle. More applicable to data with cyclical movements.

Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}...$$

Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}....$$

Methods

The following two methods are followed in moving averages,

a) Odd Yearly Method

i) Calculate 3/5...yearly totals

ii) Now compute 3/5 yearly average by dividing the totals calculated in step (i) by the respective number of years, i 3/5/...

iii) Short term oscillations are calculated using the formula, $Y - Y_C$

Where, Y - Actual value and Y - Estimated value.

b) Even Yearly Method

Example : 4 years

i) Calculate 4 yearly moving totals and place at the centre of middle two years of the four years considered.

ii) Divide 4 yearly moving totals by 4 to get 4 yearly average.

iii) Take a 2 period moving average of the moving average which gives the 4 yearly moving average centered.

10. Define least square method.

Ans :

Least square method is the most widely used method and provides us with a mathematical device to obtain an objective fit to the trend of a given time series. This method is so called because a trend line computed by this method is such that the sum of the squares of the deviation between the original data and the corresponding computed trend values is minimum. This method can be used to fit either a straight line trend or a parabolic trend.

The straight line trend equation is in the form of $Y = a + bX$

Where, Y denotes the trend value of the dependent variable

X denotes the independent variable.

a and b are constants.

The values of a and b are obtained by solving the following normal equations.

$$\Sigma Y = Na + b\Sigma X$$

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

Where, N represents the number of years in the series.

When $\Sigma X = 0$ the above normal equations are simplified to

$$a = \frac{\sum Y}{N}$$

$$b = \frac{\sum XY}{\sum X^2}$$

By substituting a and b values in straight line trend equation $Y = a + bX$, we get the straight line equation which can be used for estimation of future values.

11. What do you mean by Deseasonalisation of Data?

Ans :

One of the objectives of calculating seasonal indices is to eliminate the impact of seasonal factors on the given time series values, the process of elimination of effect of seasonal variations from the given data is called 'Depersonalization. 'Depersonalization helps in proper interpretation of data. The process depersonalization is dependent on whether Additive model or Multiplicative model is applicable to the given data.

In an additive model where $Y = T + S + C + I$, Deseasonalise Values = $Y - S = T + C + I$. In a multiplicative model where $Y = T \times S \times C \times I$.

Deseasonalised Values = $Y/S = T \times C \times I$.

12. Explain the utility of time series.

Ans :

"Time series analysis has wide spread applications in the areas of business and economics. It is also used in various other disciplines such as natural, social and physical sciences. The utility of time series is as under:

i) It helps in understanding past behavior of the given phenomenon.

ii) It helps in planning future operations and in the formulation of executive and policy decisions.

iii) It helps in evaluation of current achievement by comparing actual achievement with expected achievement.

iv) It helps in forecasting the behavior of the phenomenon in future.

v) It helps compare change in values of different phenomenon at different times.

vi) It helps in isolation of the impact of various factors affecting the time series.

Exercise Problems

1. Calculate the 3 yearly moving averages to determine trend component of the following time series.

Year :	1973	1974	1975	1976	1977	1978	1979
Annual sales :	2	6	1	5	3	7	2

2. Using three year moving averages, determine the trend and short term fluctuations.
Plot the original and trend values on the same graph paper:

Year :	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
Production:	21	22	23	25	24	22	25	26	27	26

3. Calculate the trend values by the method of least squares from the data given below and estimate the sales for the year 1985.

Year:	1976	1977	1978	1979	1980
Sales of T.V. ('000):	9	12	14	16	20

(Ans: Yc = 20 + 5 x, 44.5)

4. Fit a straight line to the following data and compute the trend values

Year:	1960	1961	1962	1963	1964	1965	1966
Sales:	3	7	6	8	9	7	10

(tons in thousands)

(Ans : Yc = 7.14 + 0.85 x)

5. Fit a straight line trend by method of least squares to the following data:

Year:	1981	1982	1983	1984	1985	1986	1987
Profit (Rs.'000):	57	65	63	72	69	78	82

(Ans: Yc = 69.43 + 3.82 x)

6. Fit a straight line to the following data and compute the trend values.

Year:	1991	1993	1994	1995	1996	1997	2000
Profits (Rs. Lakhs):	79	88	91	92	90	94	96

(Ans: Yc = 90 + 3.07 x)

Choose the Correct Answer

1. The first step in time-series analysis is to [c]
   a) Perform preliminary regression calculations
   b) Calculate a moving average
   c) Plot the data on a graph
   d) Identify relevant correlated variables

2. Time-series analysis is based on the assumption that [a]
   a) Random error terms are normally distributed
   b) There are dependable correlations between the variable to be forecast and other independent variables
   c) Past patterns in the variable to be forecast will continue unchanged into the future
   d) The data do not exhibit a trend

3. The cyclical component of time-series data is usually estimated using [d]
   a) Linear regression analysis
   b) Moving averages
   c) Exponential smoothing
   d) Qualitative methods

4. In time-series analysis, which source of variation can be estimated by the ratio-to-trend method?
   a) Cyclical
   b) Trend [c]
   c) Seasonal
   d) Irregular

5. If regression analysis is used to estimate the linear relationship between the natural logarithm of the variable to be forecast and time, then the slope estimate is equal to [c]
   a) The linear trend
   b) The natural logarithm of the rate of growth
   c) The natural logarithm of one plus the rate of growth
   d) The natural logarithm of the square root of the rate of growth

6. The use of a smoothing technique is appropriate when [a]
   a) Random behavior is the primary source of variation
   b) Seasonality is present
   c) Data exhibit a strong trend
   d) All of the above are correct

7. The greatest smoothing effect is obtained by using [b]
   a) A moving average based on a small number of periods
   b) Exponential smoothing with a small weight value
   c) The root-mean-square error
   d) The barometric method

8. The root-mean-square error is a measure of [d]
   a) Sample size.
   b) Moving average periods
   c) Exponential smoothing
   d) Forecast accuracy

9. Barometric methods are used to forecast [c]
   a) Seasonal variation b) Secular trend
   c) Cyclical variation d) Irregular variation

10. If 3 of the leading indicators move up, 2 move down, and the remaining 6 are constant, then the diffusion index is [b]
    a) 3/6 = 50% b) 3/11 = 27%
    c) 5/11 = 45% d) 6/11 = 55%

Fill in the blanks

1. A single-equation econometric model of the demand for a product is a __________ equation in which the quantity demanded of the product is an __________ variable.
2. Trend projection is an example of __________
3. Turning points in the level of economic activity can be forecast by using __________
4. The graph of time series is called__________
5. Secular trend can be classified in to __________
6. In time series seasonal variations can occur within a period of __________
7. Time-series analysis is based on the assumption that __________
8. A single-equation econometric model of the demand for a product is a __________ equation in which the quantity demanded of the product is an __________ variable
9. Barometric methods are used to forecast __________
10. A __________ is a statistical data that are collected, observed or recorded at regular intervals of time.

ANSWERS

1. Structural, endogenous
2. Time-series
3. Barometric methods
4. Histogram
5. Four methods
6. One year
7. Past patterns in the variable to be forecast will continue unchanged into the future
8. Structural, endogenous
9. Cyclical variation
10. Time series

UNIT IV	**PROBABILITY :** Probability - Meaning - Experiment - Event - Mutually Exclusive Events - Collectively Exhaustive Events - Independent Events - Simple and Compound Events - Basics of Set Theory - Permutation- Combination - Approaches to Probability: Classical - Empirical - Subjective - Axiomatic - Theorems of Probability: Addition - Multiplication - Baye's Theorem.

4.1 Probability

4.1.1 Meaning

Q1. Define the term probability?

(OR)

What is probability?

Ans :

Introduction

An Italian mathematician, Galileo (1564 - 1642), attempted a quantitative measure of probalmhjfwhile dealing with some problems related to gambling. In the middle of 17th Century, two French mathematicians, Pascal and Fermat, laid down the first foundation of the mathematical theory of probability while solving the famous 'Problem of Points' posed by Chevalier-De-Mere. Other mathematicians from several countries also contributed in no small measure to the theory of probability. Outstanding of them were two Russian mathematicians, A. Kintchine and A. Kolmogoroff, who axiomised the calculus of probability.

If an experiment is repeated under similar and homogeneous conditions, we generally come across two types of situations.

(i) The net result, what is generally known as 'outcome' is unique or certain.

(ii) The net result is not unique but may be one of the several possible outcomes.

The situations covered by :

(i) are known as 'deterministic' or 'predectable' and situations covered by

(ii) are known as 'probabilistic' or 'unpredictable'.

'Deterministic' means the result can be predicted with certainty. For example, if r is the radius of the sphere then its volume is given by $V = \frac{4}{3}\pi r^3$ which gives uniquely the volume of the sphere.

There are some situations which do not lend themselves to the deterministic approach and they are known as 'Probabilistic'.

For example, by looking at the sky, one is not sure whether the rain comes or not.

In such cases we talk of chances or probability which can be taken as a quantitative measure of certainity.

Definitions

In a random experiment, let there be n mutually exclusive and equally likely elementary events. Let E be an event of the experiment. If m elementary events form event E (are favourable to E), then the probability of E (Probability of happening of E or chance of E), is defined as

$$P(E) = \frac{m}{n} = \frac{\text{Number of elementary events in E}}{\text{Total number of elementary events in the random experiment}}$$

If $\bar{E}$ denotes the event of non-occurrence of E, then the number of elementary events in E is n-m and hence the probability of E (non-occurrence of E) is

$$P(E) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(E) \Rightarrow P(E) + P(\bar{E}) = 1$$

Since m is a non-negative integer, n is a positive integer and $m \le n$, we have m

$$0 \le \frac{m}{n} \le 1.$$

Hence $0 \le P(E) \le 1$ and $0 \le P(\bar{E}) \le 1$.

1. Statistical Probability

Suppose an experiment is repeated 'w' times under essentially identical conditions.

Let an event A happens w-times then $\frac{m}{n}$ is defined as the relative frequency of A.

Statistical probability is also known as 'relative frequency' probability. The limit of this relative frequency as $n \to \infty$ is defined as the probability of A.

$$\boxed{\therefore\ P(A) = \lim_{n\to\infty}\frac{m}{n}}\quad 0 \le \frac{m}{n} \le 1$$

2. Axiomatic Probability

In axiomatic probability three axioms or postulates are explained on the basis of which probability is calculated. They are,

(i) Probability of an event ranges from, zero to one $\Rightarrow 0 \le P(A) \le 1$

(ii) Probability of entire sample space, $P(S) = 1$

(iii) If 'A' and 'B' are mutually exclusive events, then the probability of occurrence of either A or B is $P(A \cup B) = P\{A) + P(B)$.

In general, there are three chances which can be expected for any events.

Q2. Explain the importance of probability.

Ans :

(i) The probability theory is very much helpful for making prediction. Estimates and predictions form an important part of research investigation. With the help of statistical methods, we make estimates for the further analysis. Thus, statistical methods are largely dependent on the theory of probability.

(ii) It has also immense importance in decision making.

(iii) It is concerned with the planning and controlling and with the occurrence of accidents of all kinds.

(iv) It is one of the inseparable tools for all types of formal studies that involve uncertainty.

(v) The concept of probability is not only applied in business and commercial lines, rather than it is also applied to all scientific investigation and everyday life.

(vi) Before knowing statistical decision procedures one must have to know about the theory of probability.

(vii) The characteristics of the Normal Probability. Curve is based upon the theory of probability.

4.1.2 Experiment - Event - Mutually Exclusive Events - Collectively Exhaustive Events - Independent Events - Simple and Compound Events

Q3. Explain the various terms used in probability theory.

Ans : **(Jan.-21, June-18, Imp.)**

(i) Random Experiment

If an ‘experiment’ is conducted, any number of times, under essentially identrfcal conditions, there is a set of all possible outcomes associated with it. If the result is not certain and is anyone of the several possible outcomes, the experiment is called a random trial or a random experiment. The outcomes are known as elementary events and a set of outcomes is an event. Thus an elementary event is also an event.

(ii) Outcome

The result of random experiment is usually referred as an outcome.

(iii) Event

An event is possible outcome of an experiment or a result of trial.

(a) Simple Event: In case of simple events we consider the probability of the happening or not happening of single events. For example, we might be interested in finding out the probability of drawing a red ball from a bag containing 10 white and 6 red balls.

(b) Compound Events: Compound events we consider the joint occurrence of two or more events. For example, if a bag contains 10 white and 6 red balls and if two successive draws of 3 balls are made, we shall be finding out the probability of getting 3 white balls in the first draw and 3 black balls in the second draw we are thus dealing with a compound event.

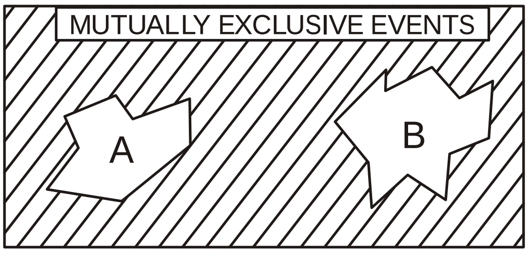
(iv) Mutually Exclusive Events

Two events are said to be mutually exclusive or incompatible when both cannot happen simultaneously in a single trial or, in other words, the occurrence of any one of them precludes the occurrence of the other.

For example, if a single coin is tossed either head can be up or tail can be up, both cannot be up at the same time. Similarly, a person may be either alive or dead at a point of time he cannot be both alive as well as dead at the same time.

To take another example, if we toss a dice and observe 3, we cannot expect 5 also in the same toss of dice. Symbolically, if A and B are mutually exclusive events, P (AB) = 0.

The following diagram will clearly illustrate the meaning of mutually exclusive events :



Disjoint Sets

It may be pointed out that mutually exclusive events can always be connected by the words "either ..... or". Events A, B, C are mutually exclusive only if either A or B or C can occur.

(v) Collectively Exhaustive Events

Events are said to be exhaustive when their totality includes all the possible outcomes of a random experiment. For example, while tossing a dice, the possible outcomes are 1, 2, 3, 4, 5 and 6 and hence the exhaustive number of cases is 6. If two dice are thrown once, the possible outcomes are :

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1,6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

The sample space of the experiment i *e..* 36 ordered pairs (6^2). Similarly, for a throw of 3 dice exhaustive number of cases will be 216 (i.e. 6^3) and for n dice they will be 6^n.

Similarly, black and red cards are examples of collectively exhaustive events in a draw from a pack of cards.

(vi) Equally Likely Events

Events are said to be equally likely when one does not occur more often than the others. For example, if an unbiased coin or dice is thrown, each face may be expected to be observed approximately the same number of times in the long run. Similarly, the cards of a pack of playing cards are so closely alike that we expect each card to appear equally often when a large number of drawings are made with replacement. However, if the coin or the dice is biased we should not expect each face to appear exactly the same number of times.

(vii) Independent Event

Two or more events are said to be independent when the outcome of one does not affect, and is not affected by the other.

For example, if a coin is tossed twice, the result of the second throw would in no way be affected by the result of the first throw. Similarly, the results obtained by throwing a dice are independent of the results obtained by drawing an ace from a pack of cards.

To consider two events that are not independent, let A stand for a firm's spending a large amount of money on advertisement and *B* for its showing an increase in sales. Of course, advertising does not guarantee higher sales, but the probability that the firm will show an increase in sales will be higher if A has taken place.

(viii) Dependent Event

Dependent events are those in which the occurrence or non-occurrence of one event in any one trial affects the probability of other events in other trials. For example, if a card is drawn from a pack of playing cards and is not replaced, this will alter the probability that the second card drawn is, say an ace. Similarly, the probability of drawing a queen from a pack of 52 cards is $\frac{4}{52}$ or $\frac{1}{13}$. But if the card drawn (queen) is not replaced in the pack, the probability of drawing again a queen is 3/51 (\ the pack now contains only cards out of which there are 3 queens).

(xi) Non-mutually Exclusive Events

When two events can occur simultaneously in a single trial then such events are said to be non-mutually exclusive events.

Eaxample, from a pack of cards, drawing a red card and drawing a queen are the two events. These two events can occur simultaneously while drawing a red queen.

Hence, these two events are said to be non-mutually exclusive events which can occur at the same time.

PROBLEMS

1. What is the probability for a leap year to have 52 Mondays and 53 Sundays?

Sol :

A leap year has 366 days i.e., 52 weeks and 2 days.

These two days can be any one of the following 7 ways :

(i) Mon & Tue

(ii) Tues & Wed

(iii) Wed & Thurs

(iv) Thurs & Fri

(v) Fri & Sat

(vi) Sat & Sun

(vii) Sun & Mon

Let E be the event of having 52 Mondays and 53 Sundays in the year.

Total number of possible cases is n = 7

Number of favourable cases to E is m = 1

(Sat & Sun is the only favourable case)

$\therefore \quad P(E) = \frac{m}{n} = \frac{1}{7}$

2. Five digit numbers are formed with 0, 1,2, 3,4 (not allowing a digit being repeated in any number). Find the probability of getting 2 in the ten's place and 0 in the units place always.

Sol :

Total number of 5 digit numbers using the digits 0, 1, 2, 3, 4 is

$= n = 4 \times 4 \times 3 \times 2 \times 1 = 96$ (or) $5! - 4! = 96$

Let E be the event of getting a number having 2 in 10's place and 0 in the units place.

So the number of numbers favourable to is = m = 3.2.1.1.1 = 6

$\therefore \quad P(E) = \frac{m}{n} = \frac{6}{96} = \frac{1}{16}$

3. In a class there are 10 boys and 5 girls. A committee of 4 students is to be selected from the class. Find the probability for the committee to contain at least 3 girls.

Sol :

A committee of 4 students out of 15 can be formed in $^{15}C_4$ ways i.e., n = $^{15}C_4$

Let E be the event of forming a committee with at least 3 girls.

Now the committee can have 1 boy, 3 girls or no boy, 4 girls. So the number of ways of forming the committee = The number of favourable ways to E

$$= {}^{10}C_1 \times {}^{5}C_3 + {}^{10}C_0 \times {}^{5}C_4 = 100 + 5 = 105$$

$$\therefore \quad P(E) = \frac{m}{n} = \frac{105}{{}^{15}C_4} = 0.0769$$

4. When two dice are thrown, find the probability that the sum of the numbers is either 10 or 11.

Sol : **(June-18)**

When 2 dice are thrown sample spaces $6^2 = 36$

The no. of of possible outcomes

$$(4, 6)\ (5, 5)\ (5, 6)\ (6, 4)\ (6, 5) = \frac{5}{36}$$

Let "A" be the Event that number selected would be sum is 10.

'B' be the Even that number selected would be sum is 11.

4.2 Basics of Set Theory

Q4. Define set. What are the different ways of representing a set?

Ans :

A set is a well-defined collection of all possible objects according to a well defined rule. The objects comprising a set are called elements or members of the set. A subset, B of a set A is another set whose element are also elements of the set A and is written as $B \subset A$. The set B is a proper subset of A if B is a subset of A and at least one element of A is not contained in B.

Two sets A and B are said to be equal or identical if $A \subset B$ and $B \subset A$, i.e., A = B or B = A.

If a set is having no elements at all, it is said to be an empty set or null set and is usually denoted by ϕ. Thus, empty set is a subset of every set. The universal set, S is the set of all elements considered in a given problem.

If there are three sets A, B and C in such a way that $A \subset B$ and $B \subset C$, then $A \subset C$. This property is called transitivity.

A set is said to be finite if it is empty or contains n elements (n > 0) otherwise it is infinite.

An infinite set is countable if its elements can be indexed by the positive integers, i.e., placed in one-to-one correspondence with them. Otherwise, it is non-countable set basic operations on sets:

(i) Union

The sum or union of two sets A and B (written as A + B or $A \cup B$) is the set of all the elements that belong to A or B or both.

$A \cup B = B \cup A \Rightarrow$ union operation is cumulative

$(A \cup B) \cup C = A \cup (B \cup C) \Rightarrow$ union operation is associative

$A \cup A = A \Rightarrow$ Idempotency law

$A \cup f = A$

(ii) Intersection

The product or intersection of two sets A and B (written as AB or $A \cap B$) is a set consisting of all elements that are common to both the sets A and B.

$A \cap B = B \cap A \Rightarrow$ commutative

$(A \cap B) \cap C = A \cap (B \cap C) \Rightarrow$ associative

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \Rightarrow$ distributive

$A \cap A = A \Rightarrow$ Idempotency law

$A \cap \phi = \phi$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(iii) Difference

The difference of A and B or the relative complement of B in A (written as A – B) is the set of all the elements belonging to A that do not belong to B.

$A - B = A \cap B'$

$= A - (A \cap B')$

For example, if A = {1, 2, 3, 4} and B = {2, 4, 6, 8}, then A – B = {1, 3} and B – A = {6, 8}. Thus $A - B \neq B - A$ and hence the difference operation is not commutative.

(iv) Symmetric Difference

The symmetric difference of two sets A and B is defined as the union of the two relative complements, i.e.,

$A \Delta B = (A - B) \cup (B - A)$

We also have A D A = ϕ

A D B = B D A

$A \cup B$ = (A D B) D ($A \cap B$)

$A \cap B$(B D C) = ($A \cap B$) D ($A \cap C$)

(v) Complement

Consider a set S and its subset A the complement or negative of A(written A' or $\bar{A}$) is the set containing all the elements of the set S that are not elements of A, i.e., A' = S – A.

A + A' = S

$A \cap A' = \phi$

(A')' = A $\Rightarrow$ Involution law

(vi) Mutually Exclusive

Two set A and B are said to be disjoint or mutually exclusive if they do not have any common elements, i.e., $A \cap B = \phi$.

(vii) De Morgan's Laws

De Morgan's laws state that the complement of a union of two sets A and B equals the intersection of the complements A' and B', i.e.,

$$(A \cup B)' = A' \cap B'$$

and the complement of intersection of two sets A and B equals the union of the complements A' and B' i.e.,

$$(A \cap B)' = A' \cup B'$$

(viii) Duality

The duality principle states that if in an identity, all the unions are replaced by intersections, all intersections by unions, S by f and f S, then the identity is preserved.

For example, in the identity,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

if unions are replaced by intersections and intersections are replaced by unions, the identity.

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

results and it is also a valid identity.

(ix) Set Theory

A set is a collection or aggregate od definite and distinguishable objects selected by means of some rules or description.

A set is a well defined collection of objects, referred to as elements or members of the set.

A set is denoted by capital letters A, B, C.. etc. and lower case letters a, b, c, ... etc., are used to denote the elements of the set.

In defining a set, "well, defined" means it is possible to decide whether a given element belongs to the collected group or not.

The statement "x is an element of Set A" or "x belongs to Set A" is written as $x \in A$.

The statement "x is not an element of A" is written as $x \notin A$.

(x) Representation of a Set

There are two ways to represent a set.

(a) Roster or Tabular From

In this form, all the elements of the set are listed and separated by commas. The elements are enclosed within braces.

Ex: Set of decimal digits

$$A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

The order of elements in a set has no significance.

Thus {0, 1, 2}, {2, 1, 0} are the representation of same set.

Repeated elements in the set can be represented only once. i.e.,

$$\{1, 3, 2, 2, 1\} = \{1, 2, 3\}$$

(b) Rule Method or Set Builder Form

In this method, a set is defined by specifying a common property possessed by the elements of the set, in common.

$A = \{x : P(x)\}$

A vertical Bar can also be used in the place of colon(:)

Example : The set A = {1, 4, 9, 16, 25} is written as

$A = \{x : x = K^2$, where K is a Natural Number $\leq 5\}$

(xi) Finite and Infinite Sets

A set with finite number of elements is called Finite set.

A = {1, 2, 3, 4}

A set with infinite number of elements is an infinite set.

A = {Integers}

(xii) Null Set

A set which contains no elements is called Null set/Void set and is denoted as f.

A = {x is a multiple of 2, x odd}

(xiii) Singleton Set

A set containing only one element is called singleton set.

A = {a}

(xiv) Sub Set

This refers within a set,

If every element of a set A also belongs to set B, then A is called a Subset of B, denoted as $A \subseteq B$ [or $A \subset B$)

If A is not a subset of B, then atleast one element of A does not belong to B, and is written as $A \not\subset B$.

Every set is its subset i.e., $A \subseteq A$. Null set is a subset of any set A i.e., $f \subseteq A$.

If A is a subset of B and B is a subset of C, then A is subset of C.

If a set 'A' contains 'n' elements, then A will have 2^n subsets.

Let A = {a, b, c}. Then, its subsets are f, {a}, {b}, {c}, {a, b}, {a, c}, {b, c}, {a, b, c}.

If A is a subset of B, then B is referred to a superset of A, written as $B \supseteq A$.

Any subset A is said to be a proper subset of another set B, if A is subset of B, and there is atleast one element of B, that does not belong to A, i.e., $A \subseteq B$ and $A \neq B$.

A = {1, 6}, B = {1, 5, 6}

A is a proper subset of B, denoted as $A \subset B$. Two sets A and B are equal if and only if their members are same.

(xv) Universal Set

It is a Non-empty set for which any set under consideration is a subset.

Q5. Show that for any two sets A and B

(i) $\rho(A) \cup \rho(B) \subseteq \rho(A \cup B)$

(ii) $\rho(A) \cap \rho(B) \subseteq \rho(A \cap B)$

Where $\rho(X)$ is the power set of X?

Sol :

Given that,

The $\rho(A)$ and $\rho(B)$ are power sets of sets A and B respectively.

To prove,

(i) $\rho(A) \cup \rho(B) \subseteq \rho(A \cup B)$

(ii) $\rho(A) \cap \rho(B) \subseteq \rho(A \cap B)$

Proof

By definition, power set of any set is the set that contains all subsets of that set. Power set is denoted by 'ρ'.

$\rho(A)$ is set of all subsets of the set A, symbolically, we can write as,

$\{a/a \subseteq A\}$ or 2^A

So, for p(B), it can also be written as,

$\{b/b \subseteq B\}$ (or) 2^B

(i) Let us prove this first statement with the help of an example, consider any two sets A = {1,2,3}, B = {1,3,6}

$\therefore$ $\rho(A) = \{0\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$

$\therefore$ $\rho(B) = \{0\}, \{1\}, \{3\}, \{6\}, \{1, 3\}, \{1, 6\}, \{3, 6\}, \{1, 3, 6\}$

L.H.S $\Rightarrow \rho(A) \cup \rho(B)$

$= \{0\}, \{1\}, \{2\}, \{3\}, \{6\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2, 3\}, \{1, 3, 6\}, \{1, 6\}, \{3, 6\}$ ...(1)

Now, find A∪B which is equal to {1, 2, 3, 6}

R.H.S $\Rightarrow \rho(A \cup B)$

$= \{0\}, \{1\}, \{2\}, \{3\}, \{6\}, \{1,2,3\}, \{1,2,6\}, \{2,3\}, \{2,6\}, \{2,3,6\}, \{3,6\}, \{1,2,3,6\}, \{1,3,6\}, \{1,6\}, \{1,2,6\}, \{1,3\}$ ... (2)

$= 2^4 = 16$ Subsets

$\therefore$ From equations (1) and (2), we get,

$\rho(A) \cup \rho(B) \subseteq \rho(A \cup B)$

(ii) To prove the second statement let us take the above example.

$\rho(A) \cup \rho(B) = \{0\}, \{I\}, \{3\}, \{1,3\}$ ...(3)

Now, $A \cap B = \{1,3\}$

$\therefore$ $\rho(A \cap B) = \{0\}, \{1\}, \{3\}, \{1, 3\}$ ...(4)

From equations (3) and (4), we get,

$\rho(A) \cap \rho(B) \subseteq \rho(A \cap B)$

6. Show that for any two sets A and B,

(a) $A - (A \cap B) = A - B$

(b) $A = A \cap B \cup (A - B)$

Sol :

(a) Given that,

Two sets are ‘A’ and ‘5’

Require to prove, $A - (A \cap B) = A - B$

For the valuer,

$y \in A - (A \cap B)$

$\Rightarrow$ $y \in \{y \backslash y \in A \cap y \notin (A \cap B)\}$

$\Rightarrow$ $y \in A \cap \sim(y \in A \cap y \in B)$

$\Rightarrow$ $y \in A \cap (y \notin A \cup y \notin B)$

$\Rightarrow$ $(y \in A \cap y \notin A) \cup (y \in A \cap y \notin B)\}$

$\Rightarrow$ $(y \in A \cap y \notin A) \cup (y \in A \cap y \notin B)\}$

$\Rightarrow$ $y \in A \cap y \notin B$

$\therefore$ $y \in \{y \backslash y \in A \cap y \notin B)$

$\therefore$ $y \in A - B$

Hence proved.

(b) Given: two sets A and B,

Require to prove: $A = (A \cap B) \cup (A - B)$

Consider,

R.H.S $= (A \cap B) \cup (A - B)$

$= x \in (A \cap B) \cup (A - B)$

$= x \in (A \cap B) \cup x \in (A \cap B')$

$= (x \in A \cap x \in B) \cup (x \in A \cap x \in B')$

$= x \in A \cap (x \in (B \cup B)$

$= x \in A \cap x \in U$ $\quad (\because A \cup A' = U)$

$= x \in (A \cap xU)$ $\quad (\because A \cap U = A)$

$= A(L.H.S)$

Hence proved.

4.3 Permutation and Combination

Q5. Explain in detail abiout Permutation and Combination.

Ans : **(Imp.)**

(i) Permutation

Arrangement of 'n' things in a specified order is called permutation. Here all things are taken at a time.

Example

Consider the letters a, b and c. Considering all the three letters at a time, the possible permutations are abc, acb, bca, bac, cba and cab.

Arrangement or 'r' things taken at a time from 'n' things where $r \leq n$, in a specified order is called r-permutation.

Example : From the three letters a, b and c, the possible 2-permutations i.e., 2 letters taken at a time are ab, ba, ac, ca, bc, cb.

The above computation is direct computation. Now we apply the counting technique. Consider the permutations considering all the three letters at a time.

The first letter can be selected in three different ways. Following this, the second letter can be selected in two different ways. And the 3rd letter can be selected in only one way. Let the three letter word be represented as,

c	a	b

The possible ways of selecting the letters is

3	2	1

Thus, the possible No. of permutations are 3.2.1 = 6. Now, consider 2-permutation.

3	2

$\therefore$ The possible No. of permutations are 3.2 = 6

The number of permutations taking 'r' things at a time from 'n' available things is denoted as p(n, r) or ^nP_r.

In finding these permutations, the first thing can be selected in 'n' ways.

Following this, the second thing can be selected in (n – 1) ways. [i.e., [n – (2 – 1)] ways]. And the third thing can be selected in (n – 2) ways, i.e., [n – (3 – 1)] ways. Thus, the different possible r-permutations is P(n, r) = n(n – 1) (n – 2) ... $(n-\overline{r-1})$

This can be written as

$$\frac{n(n-1)(n-2)\ldots(n-\overline{r-1})(n-r)(n-\overline{r-1})\ldots 1}{(n-r)(n-\overline{r-1})(n-\overline{r+2})\ldots 1} = \frac{n!}{(n-r)!}$$

$\therefore$ P(n, r) or ^nP_n = $\dfrac{n!}{(n-r)!}$ = $\left(\dfrac{\text{Factorial } n}{\text{Factorial}(n-r)}\right)$ [$\angle n$ also a notation for factorial n]

If r = n, i.e., P(n, n) = ^nP_n = $\dfrac{n!}{(n-r)!}$ = n! [$\because$ 0! = 1]

Consider the letters a, a and b. Here two letters are alike. Then the possible different permutations taking all at a time are aab, aba, baa.

Thus, from a set of 'n' objects, of which n_1 objects alike, n_2 objects are alike ... etc., the number of different possible permutation are,

$$\frac{n!}{n_1!..n_2...n_K!}$$

In the above example, it is $\frac{3!}{2!} = 3$

This concept is referred to as permutations with repetitions.

(ii) Combination

In permutations, the order of arrangement of objects is important. But, in combinations, order is not important, but only selection of objects.

For example, abc, bca, cab are different permutations but only one combination of 3 letters a, b and c.

The possible number of combinations of n objects, taken r at a time is denoted by C(n, r) or n_C.

Consider 3 letter a, b and c. The possible combinations taking two at a time are ab, bc, ac, ca, bc, cb. Thus, the number of permutations is equal to the number of combinations multiplied by 2!

$\therefore$ P(n, r) = ^nP_r = r! ^nC_r = r! ^nC_r $\Rightarrow$ ^nC_r = $\dfrac{P(n,r)}{r!} = \dfrac{\angle n}{\angle r.\angle n-r}$

Ordered Partitions

Assume that there are 10 toys to be divided between 4 children such that one of them gets 4 and remaining each gets 2 toys.

So, the given set A(=10) should be divided into ordered partions A_1(=4), A_2(=2) and (= 2).

$\{A\} = \{A_1, A_2, A_3, A_4\}$

From 10 toys, selecting the first 4 is in $^{10}C_4$ ways, i.e., A_1 can be determined in $^{10}C_4$ ways. Following this, from the remaining 6, A_2 can be determined in $^6C_2$ ways. Following this, A_3 can be determined in $^4C_2$ ways. Following this A_4 can be determined in $^2C_2$ ways.

Thus, there are $^{10}C_4 . ^6C_4 . ^4C_2 . ^2C_2$ different ordered partitions of A into A_1 consisting of 4 toys, A_2 consisting of 2 toys, A_3 consisting of 2 toys and A_4 consisting of 2 toys.

$$^{10}C_4 . ^6C_2 . ^4C_2 . ^2C_3 = \frac{10!}{4!6!} \times \frac{6!}{2!4!} \times \frac{4!}{2!2!} \times \frac{2!}{2!0!} \times \frac{10!}{4!2!2!2!} = 18900$$

Thus, this can be generalized as $\dfrac{n!}{n_1!n_2!...n_K!}$, where

A_1 contains n_1 elements, A_2 contains n_2 elements and so on.

PROBLEMS

7. (a) In how many ways can 4 boys and 3 girls sit in a row.

(b) In how many ways can they sit in a row if the boys and girls are each to sit together.

Sol :

(a) The seven persons can sit in a row in 7! ways.

(b) There are two possibilities i.e., B_1 B_2 B_3 B_4 G1 G_2 G_3 B_1 B_2 B_3 B_4.

where B_i is the i[th] boy and G_j is the J[th] girl.

In each of the possibilities. B1 can sit in 4 different ways. Following this, B_2 cans it in 3 different ways. Following this, B_3 can sit in 2 different ways. Following this, B_4 can sit in only one way.

Thus, the boys together can sit in

$4 \times 3 \times 2 \times 1 = 24$ ways

Similarly, the girls can sit together in $3 \times 2 \times 1 = 6$ ways.

For each of the possibilities, boys and girls can each sit together in $6 \times 24 = 144$ ways.

In total, there are $2 \times 144 = 288$ ways.

8. (a) How many car number plates can be made if each plate contains 2 different letters followed by 3 different digits.

(b) Solve the problem if the first digit cannot be zero.

Sol :

(a)

Letter_1	Letter_2	Digit_1	Digit_2	Digit_3

There are 26 letters. Letter 1 can be selected in 26 ways and letter 2 can be selected in 2.5 ways.

Digit 1 can be selected from 0 to 9 digits in 10 ways.

Digit 2 can be selected in 9 ways.

Digit 3 can be selected in 8 ways.

The total no. of number plates $= 26 \times 25 \times 10 \times 9 \times 8 = 4{,}68{,}000$

(b) Since 1[st] digit $\neq 0$, it can be selected in 9 ways. 2[nd] digit can be selected in 9 ways. 3[rd] digit can be selected in 8 ways.

The total no. of Number plates $= 26 \times 25 \times 9 \times 8 = 421200$

9. (a) Find the number of distinct permutation that can be formed from all of the letters of the word ELEVEN?

(b) How many of them begin and End with E?

(c) How many of them have 3E's together?

(d) How many begin with E and end with N.

Sol :

(a) In the word Eleven, three are alike. This is the case of permutation with repetition with 3 objects alike. So, the possible No. of distinct permutations.

$$= \frac{6!}{3!1!1!1!} = 120$$

(b) EXXXXE

$\Rightarrow$ In between two E's we have 4 places. From the remaining 4 letters, first place can be occupied in 4 ways, 2nd place in 3 ways, 3 rd place in 2 ways, and 4th place in 1 way.

So, the possible No. of required permutations is 4.3.2.1 = 24

(c) The possibilities are XXXEEE, XXEEEX, XEEEXX, EEEXXX.

In each of the possibilities 3! = 6 ways are there to arrange remaining 3 letters. In total, there are 4 × 6 = 24 permutations.

d) EXXXXN

The remaining 4 places are to be filled with 4 letters L, E, V, E. This is the case with 4 letters, with 2 identical. The possible No. of permutations is

$$\frac{4!}{2!1!1!} = 12$$

---

10. A student is to answer 10 out 13 questions in an examination.

(a) How many choices has he?

(b) How many if he must answer the first two questions?

(c) How many if he must answer the first or second question, but not both?

(d) How many if he must answer exactly 3 of the first 5 questions?

(e) How many if he must answer atleast 3 of the first 5 questions?

Sol :

a) The possible no. of choices = $^{13}C_{10}$

$$= \frac{13!}{10!3!} = 286$$

b) Since first two questions must be answered, the student has to answer 8 from 11 questions. The No. of choices = $^{11}C_8$ = 165

c) If the first question is answered, excluding 2nd question, he was to answer 9 questions from 11 remaining questions. Same is the case, if second question is answered i.e., then first question is to be excluded.

$\therefore$ The No. of choices = 2 . $^{11}C_9$ = 110

d) If exactly 3 of the first five are to be answered, he has to answer 7 questions from the remaining 8.

3 questions from 5 and 7 questions from 8 results in $^5C_3 \cdot {}^8C_7$ choices = 80

e) Atleast 3 of the first 5 questions means, the possible combinations are

$^5C_3 \cdot {}^8C_7 + {}^5C_4 \cdot {}^8C_6 + {}^5C_5 \cdot {}^8C_5 = 276$

11. Show that the chances of throwing six with 4, 3 or 2 dice respectively are as 1:6:18.

Sol :

With four dice the possible combinations to get 6 are

1122, 1131, 1113, 1212, 1221, 1311, 2112, 2121, 2211, 3111

i.e., 10 combinations are there

So, the probability for getting six with four dice $= \frac{10}{6 \times 6 \times 6 \times 6}$

With 3 dice the possible combinations to get 6 are

114, 141, 123, 132, 222, 231, 213, 321, 312, 312, 411

i.e., 10 combinations are there

So, the probability for getting six with 3 dice $= \frac{10}{6 \times 6 \times 6}$

With 2 dice, the possible combinations to get 6 are 15, 51, 24, 42, 33

i.e., 5 combinations are there

So, the probability for getting six with 2 dice $= \frac{5}{6 \times 6}$

Required in the given problem is $\frac{10}{6 \times 6 \times 6 \times 6} : \frac{10}{6 \times 6 \times 6} : \frac{5}{6 \times 6}$

12. A fair coin is tossed 4 times. Find the probability that there will appear.

(a) 2 heads

b) 1 tail and 3 heads

c) at least one head

d) exactly one head

e) Not more than one head

Sol :

The distribution is a binomial distribution.

Here n = 4

Let the probability of getting a head in one toss = p

Since, the coin is fair P = 1/2

i.e., chance of success = p and chance of failure = q = 1 – p

The probability of getting 'm' number of success in 'n' number of tosses is

$P(X = m) = {}^{n}C_{m} \cdot p^{m} \cdot q^{n-m}$

(a) $P(X = 2) = {}^4C_2 . \left(\frac{1}{2}\right)^2 . \left(\frac{1}{2}\right)^{4-2} = \frac{1\times2\times3\times4}{1\times2\times1\times4} . \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$

OR

The possible number of outcome = 16, each with a probability of 1/16. There are 6 possible outcomes to get 2 heads. So, the required probability is $\frac{6}{16}$ i.e., $\frac{3}{8}$.

(b) 3 heads means 3 successes and 1 failure

$\therefore$ $P(X = 3) = {}^4C_3 . \left(\frac{1}{2}\right)^3 . \left(\frac{1}{2}\right)^{4-3} = \frac{1\times2\times3\times4}{1\times2\times3\times4} . \frac{1}{8} . \frac{1}{2} = \frac{1}{4}$

OR

There are 4 possible outcomes to get 3 heads and one tail. The corresponding probability is

$\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{4}{16} = \frac{1}{4}$

(c) at least one head

$\Rightarrow$ $P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1 - P(X = 0)$

$= 1 - {}^4C_0 . \left(\frac{1}{2}\right)^0 . \left(\frac{1}{2}\right)^4 = 1 - \frac{1}{16} = \frac{15}{16}$

(d) Exactly one head $\Rightarrow$ 1 success, 3 failures

$P(X = 1) = {}^4C_1 . \left(\frac{1}{2}\right)^1 . \left(\frac{1}{2}\right)^{4-1} = \frac{1}{4}$

OR

There are four possible outcomes to get exactly one head.

$\therefore$ $P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1 - P(X = 0)$

$= 1 - {}^4C_0 . \left(\frac{1}{2}\right)^0 . \left(\frac{1}{2}\right)^4 = 1 - \frac{1}{16} = \frac{15}{16}$

(d) Exactly one head $\Rightarrow$ 1 success, 3 failures

$P(X = 1) = {}^4C_1 . \left(\frac{1}{2}\right)^1 . \left(\frac{1}{2}\right)^{4-1} = \frac{1}{4}$

OR

There are four possible outcomes to get exactly one head.

$\therefore$ $P(X = 1) = 4. \frac{1}{16} = \frac{1}{4}$

(e) Not more than one head $\Rightarrow$ $P(X = 0) + P(X = 1)$

$= {}^4C_0 . \left(\frac{1}{2}\right)^0 . \left(\frac{1}{2}\right)^{4-0} + {}^4C_1 . \left(\frac{1}{2}\right)^1 . \left(\frac{1}{2}\right)^{4-1} = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$

4.4 Approaches to Probability

Q6. What are the approaches of probability.

Ans :

They are four different approaches on Broadly. The concept of probability. They are ass follows.

1. Classical (or) priori probability
2. Relative/empirical probability
3. Subjective approach
4. Axiomatic approach

4.4.1 Classical

Q7. Explain in detail about Classical approach used in probability.

Ans :

The probability of a given event is an expression of likelihood or chance of occurrence or an event. A probability is a number which ranges from 0 (zero) to (one) - zero for an event which cannot occur and 1 for an event certain to occur. How the number is assigned would depend on the interpretation of the term 'probability'. There is no general agreement about its interpretation and many people associate probability and chance with nebulous and mystic ideas. However, broadly speaking there are four different schools of thought on the concept of probability.

If after 'n' repetitions of an experiment, where n is very large, an event is observed to occur in K of these, then the probability of the event is K/n.

Consider an experiment which has 'n' exhaustive, mutually exclusive and equally likely events with 'K' events from the 'n' in favour of happening of the event A, then the probability 'P' of A is

$$\boxed{P(A) = \frac{K}{n}}$$

$$= \frac{\text{No. of events in favour of A}}{\text{No. of exhaustive equally likely and mutually exclusive events of the experiment}}$$

P(A) is also known as probability of success.

Limitations

Classical theory does not holds good,

(a) When all the outcomes are not equally likely

(b) When the collectively exhaustive events of an experiment are infinite

(c) Classical theory does not provide answers to certain question which occurs in our daily life.

For example,

What is the probability of occurrence of rain now? The chances of bulb getting failed etc.

4.4.2 Empirical

Q8. Explain in detail about Empirical approach used in probability.

Ans :

In the 1800s, British statisticians, interested in a theoretical foundation for calculating risk of losses in life insurance and commercial insurance, began defining probabilities from statistical data collected on births and deaths. Today this approach is called relative frequency of occurrence.

This classical definition is difficult or impossible to apply as soon as we deviate from the fields of coins, dice, cards and other simple games of chance. Secondly, the classical approach may not explain actual results in certain cases.

For example, if a coin is tossed 10 times we may get 6 heads and 4 tails. The probability of a head is thus 0.6 and that of a tail 0.4. However, if the experiment is carried out a large number of times we should expect approximately equal number of heads and tails. As n increases, i.e., approaches ∞ (infinity), we find that the probability of getting a head or tail approaches 0.5. The probability of an event can thus be defined as the relative frequency with which it occurs in an indefinitely large number of trials. If an event occurs a times out of n, its relative frequency is $\frac{a}{n}$ the value which is approached by $\frac{a}{n}$ when n becomes infinity is called the limit of the relative frequency.

Symbolically $P(A) = \underset{n\to\infty}{Lt} \frac{a}{n}$

Theoretically, we can never obtain the probability of an event as given by the above limit. However, in practice we can only try to have a close estimate of P(A) based on a large number of observations. i.e., n. For practical convenience, the estimate of P(A) can be written as if it were actually P(A) and the relative frequency definition of probability may be expressed as:

$$P(A) = \frac{a}{n}$$

In the relative frequency definition the fact that the probability is the value which is approached by $\frac{a}{n}$ when n becomes infinity, emphasises a very important point, i.e., probability involves a long- term concept. This means that if we toss a coin only 10 times, we may not get exactly 5 heads and 5 tails. However, as the experiment is carried out larger and larger number of times, say, coin is thrown 10,000 times, we can expect heads and tails very close to 50 per cent.

The two approaches, classical and empirical, though seemingly same, differ widely. In the former, P(A) and $\frac{a}{n}$ were practically equal when n was large whereas in the latter we say that P(A) is the limit $\frac{a}{n}$ as n tends to infinity. In the second approach, thus, the probability itself is the limit of the relative frequency as the number of observations increases indefinitely.

Limitations

(i) As the experiments are repeated large number of times, it takes large amount of time.

(ii) During the experimental time, the conditions may not be always identical and homogenous.

4.4.3 Subjective

Q9. Explain in detail about Subjective approach used in probability.

Ans :

The subjective approach to assigning probabilities was introduced in the year 1926 by Frank Ramsey in his book, The Foundation of Mathematics and other Logical Essays. The concept was further developed by Bernard Koopman, Richard Good and Leonard Savage. The subjective probability is defined as the probability assigned to an event by an individual based on whatever evidence is available. Hence such probabilities are based the beliefs of the person making the probability statement.

For example, if a teacher wants to find out the probability of Mr. X topping in M. Com. examination in Delhi University this year, he may assign a value between zero and one according to his degree of belief for possible occurrence. He may take into account such factors as the past academic performance, the views of his other colleagues, the attendance record, performance in periodic tests, etc., and arrive at a probability figure.

This concept emphasises the fact that since probability of an event is the degree of belief or degree of confidence placed in the occurrence of an event by a particular individual based on the evidence available to him. different individuals may differ in their degrees of confidence even when offered the same evidence. This evidence may consist of relative frequency of occurrence data and any other quantitative or non-quantitative information. Persons might arrive at different probability assignments because of differences in values, experience and attitudes, etc. If an individual believes that it is unlikely that an event will occur, he will assign a probability close to zero to its occurrence. On the other hand, if he believes that it is very likely that the event will occur, he will assign a probability close to one.

The personalistic approach is very broad and highly flexible. It permits probability assignment to events for which there may be no objective data, or for which there may be a combination of subjective and objective data. However, one has to be very careful and consistent in the assignment of these probabilities otherwise the decisions made may be misleading. Used with care the concept is extremely useful in the context of situations in business decision-making.

4.4.4 Axiomatic

Q10. Explain the Axiomatic approach to probability.

Ans : **(Jan.-21, Imp.)**

The axiomatic approach to probability was introduced by the Russian mathematician A. N. Kolmogorov in the year 1933. Kolmogorov axiomised the theory of probability and his book Foundations of Probability, published in 1933, introduces probability as a set function and is considered as a classic. When this approach is followed, no precise definition of probability is given, rather we give certain axioms or postulates on which probability calculations are based. The whole field of probability theory for finite sample spaces* is based upon the following three axioms :

1. The probability of an event ranges from zero to one. If the event cannot take place its probability shall be zero and if it is certain, i.e., bound to occur, its probability shall be one.
2. The probability of the entire sample space is 1, Le., P(S) = 1.
3. If A and B are mutually exclusive (or disjoint) events then the probability of occurrence of either A or B denoted by $P(A \cup B)$ shall be given by :

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

   It may be pointed out that out of the four interpretations of the concept of probability, each has its own merits and one may use whichever approach is convenient and appropriate for the problem under consideration.

   The probability of a event A, denoted by P(A) is so chosen as to satisfy the following three axioms.

i) $P(A) \geq 0 \Rightarrow$ This axiom states that the probability of occurrence of an event A in a random experiment may be zero or any positive number and it must not be negative number.

ii) $P(S) - 1 \Rightarrow$ This states that the sample space, S, itself is an event and since it is the event comprising all possible outcomes, it should have the highest possible probability, i.e., one.

iii) If $A \cap B = 0$, Then $P(A \cup B) = P(A) + P(B) \Rightarrow$ This axiom states that the probability of the event equal to the union of any number of mutually exclusive events is equal to the sum of the individual even probabilities.

PROBLEMS

13. A class consists of 6 girls and 10 boys. If a committee of 3 is chosen at random from the class, find the probability that (i) 3 boys are selected (ii) exactly 2 girls are selected.

Sol :

Total number of students = 16

n(S) = no. of ways of choosing 3 from 16 = $^{16}C_3$

(i) Suppose 3 boys are selected. This can be done in $^{10}C_3$ ways.

Here n(E) = $^{10}C_7$

$\therefore$ P(E) = The probability that 3 boys are selected = $\frac{n(E)}{n(S)}$

$$= \frac{^{10}C_7}{^{16}C_7} = \frac{10\times 9\times 8}{16\times 15\times 14} = \frac{3}{14} = 0.2143$$

(ii) Suppose exactly 2 girls are selected. Then

n(E) = $^{6}C_2 \times {}^{10}C_1$

$$\therefore \quad P(E) = \frac{n(E)}{n(S)} = \frac{^{6}C_2 \times {}^{10}C_1}{^{16}C_3} = \frac{15}{56} = 0.2678$$

---

14. A and B throw alternately with a pair of ordinary dice. A wins if he throws 6 before B throws 7 and B wins if he throws 7 before A throws 6. If A begins, show that his chance of winning is 30/61.

Sol :

When two dice are thrown , we have n(s) = 36

The probability of A throwing '6' = $\frac{5}{36}$ i.e., P(A) = $\frac{5}{36}$

The probability of A not throwing '6' is and is given by

$$P(\bar{A}) = 1 - P(A) = 1 - \frac{5}{36} = \frac{31}{36}$$

The probability of B throwing '7' = $\frac{6}{36}$ i.e., P(B) = $\frac{6}{36} = \frac{1}{6}$

The probability of B not throwing 7 is $P(\bar{B}) = 1 - P(B) = 1 - \frac{6}{36} = \frac{5}{6}$

$\therefore$ Chances of winning of 'A' is

$$= P(A) + P(\bar{A})P(\bar{B})P(A) + P(\bar{A})P(\bar{B})P(\bar{A})P(\bar{B})P(A) + ...$$

$$= \left(\frac{5}{36}\right) + \left(\frac{31}{36}\times\frac{5}{6}\times\frac{5}{36}\right) + \left(\frac{31}{36}\times\frac{5}{6}\times\frac{31}{36}\times\frac{5}{6}\times\frac{5}{36}\right) + \ldots\ldots$$

$$= \frac{5}{36}\left[1+\left(\frac{31}{36}\times\frac{5}{6}\right)+\left(\frac{31}{36}\times\frac{5}{6}\right)^2+\ldots\ldots\right] = \frac{5}{36}\left[\frac{1}{1-\left(\frac{31}{36}\times\frac{5}{6}\right)}\right] = \frac{30}{61}$$

4.5 Theorems of Probability

4.5.1 Addition

Q11. Explain Addition theorem of probability.

Ans :

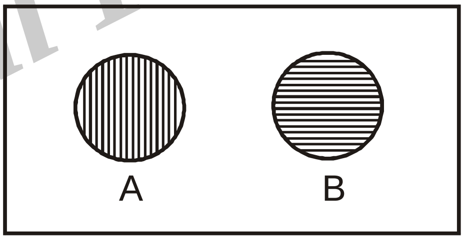
Adition theorem is different for mutually exclusive non-mutually exclusive events.

(i) For Mutually Exclusive Events

When 'A' and 'B' are two mutually exclusive events (i.e., both cannot occur at the same time) then the prooability of occurrence of A or B is equal to the sum of their individual probabilities.

$$P(A \text{ or } B) = P(A) + P(B)$$
$$\Rightarrow P(A \cup B) = P(A) + P(B)$$

Diagrametrically it can be represented as,



Mutually exclusive events

Figure: Mutually Exclusive Events

In case of 3 events A, B and C,

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

(ii) For Non-Mutually Exclusive Events

In case of non-mutually exclusive event (i.e., if the events occur together) these is a variation in the addition theorem.

When 'A' and 'B' are non-mutually exclusive events then the probability of occurrence of A or B is the sum of their individual probability which should be deducted from the probability of A and B occurring together.

$$P(A \text{ or } B) \Rightarrow P(A \cup B) = P(A)+P(B)-P(A \cap B)$$

Diagrametically it can be represented as,

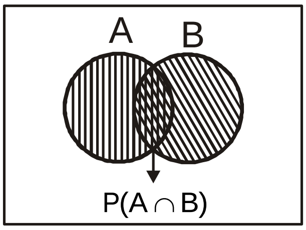


Fig.: Non-Mutually Exclusive Events

In case of three non-mutually exclusive events.

A, B and C the probability of occurence of A or B or C can be calculated by the following formula,

$$\boxed{P(A \text{ or } B \text{ or } C) = P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)}$$

PROBLEMS

15. A card is drawn from a well shuffled pack of cards. What is the probability thatdt is either a spade or an ace ?

Sol :

Let S is the sample space of all the simple events.

$\therefore \quad n(s) = 52$

Let A denote the event of getting a spade and B denote the event of getting an ace.

Then $A \cup B$ = The event of getting a spade or an ace

$A \cap B$ = The event of getting a spade and an ace

$\therefore \quad P(A) = \frac{13}{52}, P(B) = \frac{4}{52}, P(A \cap B) = \frac{1}{52}$

By Addition Theorem,

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

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

16. Three students A, B, C are in running race. A and B have the same probability of winning and each is twice as likely to win as C. Find the probability that B or C wins.

Sol :

$A \cup B \cup C = S$ = Sample space of race

By data, $P(A) = P(B)$ and $P(A) = 2\,P(C)$ .....(1)

We have $P(A) + P(B) + P(C) = 1 \Rightarrow 2P(C) + 2\,P(C) + P(C) = 1$ [by (1)]

$\Rightarrow P(C) = \frac{1}{5} P(A) = \frac{2}{5}$ and $P(B) = \frac{2}{5}$

The probability that B or C wins $= P(B \cup C)$

$= P(B) + P(Q - P(B \cap C)$

$= \frac{2}{3} + \frac{1}{5} - 0 = \frac{3}{5}$

---

17. From a city 3 news papers A, B, C are being published. A is read by 20%, B is read by 16%, C is ready by 14% both A and B are read by 8%, both A and C are read by 5% both B and C are read by 4% and all three A, B, C are read by 2%. What is the percentage of the population that read at least one paper.

Sol :

Given, $P(A) = \frac{20}{100}$, $P(B) = \frac{16}{100}$, $P(C) = \frac{14}{100}$ and

$P(A \cap B) = \frac{8}{100}$, $P(A \cap C) = \frac{5}{100}$, $P(B \cap C) = \frac{4}{100}$ and $P(A \cap B \cap C) = \frac{2}{100}$

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

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} + \frac{8}{100} - \frac{4}{100} - \frac{5}{100} + \frac{2}{100} = \frac{35}{100}$$

$\therefore$ Percentage of the population that read at least one paper $= \frac{35}{100} \times 100 = 35$

---

18. One card is drawn at random from a pack of 52 cards. What is the probability that it is either a king or a queen?

Sol : **(June-19, Imp.)**

Probability of Drawing a King Card

Let P(A) denoted as probability of drawing a king card from a pack of cards.

Total number of king cards = 4

1 kind card is drawn from 4 king cards = $4c_1 = 4$.

Let total No. of Playing cards in a pack = 52.

1 card is drawn from 52 cards = $52c_1 = 52$

$\therefore P(A) = \frac{4}{52}$

Probability of Drawing a Queen Card

Let P(B) denoted as probability of drawing 1 Queen card from a pack of cards.

Total No. of Queen cards = 4

1 card is drawn from 4 Queen cards = $4c_1 = 4$.

Let total No. of playing cards in a pack = 52

1 card is drawn from 52 cards = $52c_1 = 52$.

$\therefore$ P(B) = $\frac{4}{52}$

Probability of drawing a king or Queen is

$$P(A \cup B) = P(A) + P(B) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} \text{(or)} \frac{2}{13}.$$

19. A bag contains 4 defective and 6 good Electronic Calculators. Two calculators are drawn at random one after the other without replacement. Find the probability that

i) Two are good

ii) Two are defective and

iii) One is good and one is defective.

Sol : **(June-19, Imp.)**

Total Number of calculators in a bag = 4 + 6 = 10

(i) Probability of Drawing 2 good Calculators

Let P(A) denoted as drawing 1 good calculator from total calculators.

$\therefore$ P(A) = $\frac{6}{10}$

Let P(B) denoted as drawing 2nd good calculator without replacing the 1st calculator.

Total good calculators after first calculator is not replaced = 6 – 1 = 5.

Total Number of calculators after first calculator is drawn and not replaced = 10 – 1 = 9.

$\therefore$ P(B) = $\frac{5}{9}$

$\therefore$ Probability of drawing 2 good calculators .

$$P(A) \cdot P(B) = \frac{6}{10} \times \frac{5}{9} = \frac{30}{90} = \frac{1}{3}$$

(ii) Probability of Getting Two Defectives

Let P(A) denoted as drawing a defective calculator

Total Number of defective calculators = 4

1 Calculator is drawn from 4 = $4c_1$ = 4.

Total calculators in the bag = 4 + 6 = 10

One calculator is drawn from 10 = $10c_1$ = 10.

$\therefore$ P(A) = $\frac{4}{10}$

Let P(B) denoted as drawing another defective calculator without replacing the first.

Total number of defective calculator after first calculator is drawn and not replaced = 4 – 1 = 3.

Total number of calculators in bag after first calculator is drawn and not replaced = 10 – 1 = 9.

$\therefore$ P(B) = $\frac{3}{9}$

$\therefore$ Probability of drawing two are defective without replacing is P(A) . P(B).

$$= \frac{4}{10} \times \frac{3}{9} = \frac{12}{90} = \frac{2}{15}$$

(iii) One is good and one is defective.

Let P(A) denoted as drawing 1 good calculator.

Number of good calculators = 6.

1 calculator is drawn from 6 = $6c_1$ = 6.

Total number 9 calculators in bag = 10

1 calculator is drawn from to = $10c_1$ = 10.

$\therefore$ P(A) = $\frac{6}{10}$

Let P(B) denoted as drawing another calculator which is defective.

Total Number of defective calculators = 4.

1 calculator is drawn from 4 = $4c_1$ = 4.

Total number of calculators after first calculator is drawn and not replaced = 10 – 1 = 9.

$\therefore$ P(B) = $\frac{4}{9}$

Probability of drawing 1 good and 1 bad is P(A) . P(B)

$$= \frac{6}{10} \times \frac{4}{9} = \frac{24}{90} = \frac{4}{15}.$$

4.5.2 Multiplication

Q12. Explain Multiplication theorem of probability.

Ans :

If 'A' and 'B' are two independent events then the probability of occurrence ofboth the events is equal to the product of their individual probabilities.

For independent events,

$\boxed{P(A \cap B) = P(A) \cdot P(B)}$

Similarly,

$\boxed{P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)}$ and so on.

If 'A' and 'B' are two dependent events, in such a case multiplication theorem is altered and is given as follows. For dependent events,

$$\boxed{\begin{aligned}P(A \cap B) &= P(A / B) \cdot p(B) \\ &= P(B / A) \cdot P(A)\end{aligned}}$$

Where, P(A/B) is a conditional probability of A given that B has occurred (Thq probability of occurrence of event A when event B has already occurred is the conditional probability of A given B).

PROBLEMS

20. Find the probability of drawing 2 red balls in succession from a bag cojitdining 4 red and 5 black balls when the ball that is drawn first is (?) not replaced (ii) replaced.

Sol :

Let E_1 be the event of drawing a red ball in the first draw and £, be the event of drawing a red ball in second draw also.

(i) After the first draw the ball is not replaced. The first ball can be drawn in 9 ways and the second in 8 ways since the first ball is not replaced. Then both the balls can be drawn in 9x8 ways.

There are 4 ways in which £) can occur and 3 ways in which E_2 can occur, so that E_1 and E_2 can occur in 4 × 3 ways.

$$p\left(\frac{E_2}{E_1}\right) = P(E_2, \text{ given the probability of } E_1)$$

$$= P(\text{2nd ball is red, given that first ball is red}) = \frac{3}{8}$$

$$\therefore \quad P(E_1 \cap E_2) = P(E_1) \times P\left(\frac{E_2}{E_1}\right) = \frac{4}{9} \times \frac{3}{8} = \frac{1}{6} \quad \left[\because\ P(E_1) = \frac{4}{9}\right]$$

(ii) Suppose the ball is replaced after the first draw. Then

$$P(E_1 \cap E_2) = \frac{4}{9} \cdot \frac{4}{9} = \frac{16}{81}$$

21. A class has 10 boys and 5 girls. Three students are selected at random one after another. Find the probability that (i) first two are boys and third is girl (ii) First and third are of same sex and the second is of opposite sex.

Sol :

Total no. of students = 10 + 5 = 15

(i) The probability that first two are boys and the third is girl is

$$P(E_1 \cap E_2 \cap E_3) = \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{5}{13} = \frac{15}{91}$$

(ii) Suppose the first and third are boys and second is a girl

$$\text{Probability of the event} = P(E_1) = \frac{10}{15} \cdot \frac{5}{14} \cdot \frac{9}{13} = \frac{15}{91} = 0.1648$$

Suppose first and third are girls and second is boy.

Then the probability of the event = $P(E_2) = \frac{5}{15} \cdot \frac{10}{14} \cdot \frac{4}{13} = \frac{20}{273}$

$\therefore$ Required probability = $P(E_1) + P(E_2)$

$$= \frac{15}{91} + \frac{20}{273} = \frac{45+20}{273} = \frac{65}{273} = 0.238$$

22. Two marbles are drawn in succession from a box containing 10 red, 30 whiterf20 blue and 15 orange marbles, with replacement being made after each draw. Find the probability that

(i) Both are white

(ii) First is red and second is white.

Sol :

Total no. of marbles in the box = 75

(i) Let E_1 be the event of the first drawn marble is white. Then

$$P(E_1) = \frac{30}{75}$$

Let E_2 be the event of second drawn marble is also white. Then

$$P(E_2) = \frac{30}{75}$$

The probability that both marbles are white (with replacement)

$$= P(E_1 \cap E_2) = P(E_1).\, P(E_2 \backslash E_1) = \frac{30}{75} \cdot \frac{30}{75} = \frac{4}{25}$$

(ii) Let E_1 be the event that the first drawn marble is red. Then

$$P(E_1) = \frac{10}{75} = \frac{2}{15}$$

Let E_2 be the event that the drawn marble is white. Then

$$P(E_2 \backslash E_1) = \frac{30}{75} = \frac{2}{5}$$

$\therefore$ The probability that the First marble is red and Second marble is white

$$= P(E_1 \cap E_2) = P(E_1).\, P(E_2 \backslash E_1)$$

$$= \frac{2}{15} \cdot \frac{2}{5} = \frac{4}{75}$$

23. Three boxes, practically indistinguishable in appearance have two drawers each. Box 1 contains a gold coin in first and silver coin in the other drawer, Box 2 contains a gold coin in each drawer and Box 3 contains a silver coin in each drawer. One box is chosen at random and one of its drawers is opened at random and a gold coin is found. What is the probability that the other drawer contains a coin of silver.

Sol :

Let E denote the event that the box is choosen, i = 1,2,3.

$$P(E_1) = \frac{1}{3} \text{ for } i = 1, 2, 3$$

Let A be the event that the gold coin is chosen. Then

$P(A|E_i)$ = Probabiity that a gold coin is choosen from the box i = 1, 2, 3.

$\therefore \quad P(A|E_1) = \frac{1}{2}$ ($\because$ The total no. of coins in box 1 is 2)

$P(A|E_2) = \frac{2}{2} = 1$ (There are two gold coins in box 2)

and $P(A|E_3) = \frac{0}{2} = 0$ (There is no gold coin in box 3)

(i) The probability that the drawn coin is gold

$$P(A) = P(E_1)\,P(A|E_1) + P(E_2)\,P(A|E_2) + P(E_3)\,P(A|E_3)$$

$$= \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 1 + 0 = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

(ii) The probability that the drawn coin is silver

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

---

24. Two digits are selected at random from the digits 1 through 9.

(i) If the sum is odd, what is the probability that 2 is one of the numbers selected ?

(ii) If 2 is one of the digits selected, what is the probability that the sum is odd ?

Sol :

The given set consists of five odd digits (1, 3, 5,7, 9) and four even digits (2, 4, 6, 8).

We know that

even + even = even

even + odd = odd

odd + even = odd

odd + odd = even

(i) Total number of events = 5 × 4 = 20

If '2' is one of the digits, then the other digit must be odd.

$\therefore$ Number of ways = 5

So, required probability = $\frac{5}{20} = \frac{1}{4}$

(ii) If 2 is selected, then the remaining number of digits = 8

$\therefore$ Total events = 8

If 2 is one of the digits selected, the probability that the sum is odd

$$= \frac{\text{Favourable cases for odd}}{\text{Total events}} = \frac{5}{8}$$

25. There are 12 cards numbered 1 to 12 in a box, if two cards are selected what is the probability that sum is odd

(a) With replacement ?

(b) Without replacement ?

Sol :

The cards having even number = {2, 4, 6, 8, 10, 12}

The cards with odd numbres = {1, 3, 5, 7, 9, 11}

(a) With replacement : Suppose we select a card place it back and select one card.

This can happen in 12.12 = 144 ways.

To have an odd sum,

(i) the first card has to be odd and the second has to be even or

(ii) the first card has to be even and the second has to be odd.

Number of these favourable ways = 6.6 + 6.6 = 72 72

$\therefore$ P(sum is odd) = $\frac{22}{144} = \frac{1}{2}$

(b) Without replacement : The total number of outcomes = (12)(11) = 132

The sum can be odd as below :

(i) The card in first draw is even and on the second is odd.

Probability (for (ii)) = $\frac{6}{12} \cdot \frac{6}{11} = \frac{36}{132}$

(ii) The card in first draw is odd and on the second is even.

P(for (ii)) = $\frac{6}{12} \cdot \frac{6}{11} = \frac{36}{132}$

$\therefore$ P(the sum is odd) = $\frac{36}{132} + \frac{36}{132} = \frac{72}{132} = \frac{6}{11}$

26. A box contains 8 Red and 5 White balls. Two successive draws of 3 balls are made at random. Find the probability that the first three are white and second three are red.

(i) When there is replacement and

(ii) When there is no replacement.

Sol : **(Jan.-21, Imp.)**

(i) When there is replacement

Total No.of balls = 8 + 5 = 13

3 balls can be drawn from 13 in 13 c_3 ways

3 red balls can be drawn from 8 in 8 c_3 ways

3 white balls can be drawn form 5 in 5 c_3 ways

$$P(A \cap B) = P(A).P(B/A)$$

$$P(A) = \frac{5c_3}{13c_3} = \frac{5}{143}$$

$$P(B/A) = \frac{8c_3}{13c_3} = \frac{28}{143}$$

$$P(A \cap B) = \frac{5}{143} \times \frac{28}{143} = \frac{140}{20,449} = 0.007$$

(ii) Where there is no replacement

$$P(A \cap B) = P(A).P(B/A)$$

$$P(A) = \frac{5c_3}{13c_3} = \frac{5}{143}$$

$$P(A \cap B) = \frac{5}{143} \times \frac{7}{24} = \frac{7}{429} = 0.0102$$

27. From 30 tickets marked with first 30 numerals, 1 ticket is drawn at random. It is then replaced and a second draw is made. Find the probability that in the first draw it is multiple of 5 or 7 and in second draw it is a multiple of 3 or 7.

Sol : **(June-18, Imp.)**

n (s) = 30

Let 'A' denotes the events of drawing a multiple by 5

'B' denotes the events of drawing a multiple by 7

$$n(A) = \{5, 10,15, 20, 25, 30\} = \frac{6}{30}$$

$n(B) = \{7, 14, 21, 28\} = \frac{4}{30}$

$P(A) = \frac{n(A)}{n(S)} = \frac{6}{30}$; $P(B) = \frac{n(B)}{n(S)} = \frac{4}{30}$

i) Probability that in the first draw it is multiple by 5 or 7

$P(A) = \frac{6}{30}$; $P(B) = \frac{4}{30}$

[Mutually disjoint events)

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

$\frac{6}{30} + \frac{4}{30}$

$P(A \cup B) = \frac{10}{30}$ or 0.333

ii) Probability that in the 2nd draw it is multiple by 3 or 7

$n(A) = \{3, 6, 9, 12, 15, 18, 21, 24, 27, 30\} = 10$

$n(B) = \{7, 4, 21, 28\} = 4$

$n(A \cap B) = \{21\} = 1$

$P(A) = \frac{10}{30}$, $P(B) = \frac{4}{30}$, $P(A \cap B) = \frac{1}{30}$

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

$\frac{10}{30} + \frac{4}{30} - \frac{1}{30}$

$P(A \cup B) = \frac{13}{30}$ or 0.433

4.5.3 Baye's Theorem

Q13. State and explain Baye's probability theorem.

Ans : **(Imp.)**

$E_l, E_2, \ldots, E_n$ are n mutually exclusive and exhaustive events such that P(E.) > 0 (z = 1, 2, , ri) in a sample space S and A is any other event in S intersecting with every E. (i.e., A can only occur in combination with any one of the events E_1 $E_2, \ldots, E_n$) such that P(A) > 0.

If E_1 is any of the events of E_1 E_2, ,E„ where $P(E_t)$, $P(E_2)$,....,P(EJ and $P(A/E_1)$, $P(A/E_2)$,....,$P(A/E_1)$ are known, then

$$P(E_K \backslash A) = \frac{P(E_K).P(A/E_K)}{P(E_1).P(A/E_1) + P(E_2).P(A/E_2) + \ldots + P(E_n).P(A/E_n)}$$

Proof :

$E_1, E_2,...,E_n$ are n events of S such that $P(E_1) > 0$ and $E_i \cap E_j = \phi$ for $i \neq j$ where i, j = 1, 2,....,n. Also $E_1, E_2,....,E_n$ are exhaustive events of S and A is any other event of S where $P(A) > 0$.

$S = E_1 \cup E_2 \cup .... \cup E_N$ and

$A = A \cap S = A \cap (E_1 \cup E_2 \cup ..... \cup E_n)$

$= (A \cap E_1) \cup (A \cap E_2) \cup ..... \cup (A \cap E_n)$

Here $A \cap E_1, A \cap E_2,.....,$ are mutually exclusive events. Then

$$P(E_K/A) = \frac{P(E_K \cap A)}{P(A)} = \frac{P(E_K \cap A)}{P[(A \cap E_1) \cup (A \cap E_2) \cup ..... \cup (A \cap E_n)]}$$

$$= \frac{P(E_K \cap A)}{P(A \cap E_1) + P(A \cap E_2) + ... + P(A \cap E_n)}$$

$$= \frac{P(E_K).P(A / E_K)}{P(E_1).P(A / E_1) + P(E_2).P(A / E_2) + ... + P(E_n).P(A / E_n)}$$

Note :

Baye's theorem is also known as formula for the Probability of "Causes", i.e., probability of a particular (cause) E given that event A has happened (already).

$P(E_i)$ is'a priori probability' known even before the experiment, $P(A/E_i)$ "Likelihoods and $P(E_i / A)$ 'Posteriori Probabilities' determined after the result of the experiment.

Q14. What are the applications of Baye's Theorem?

Ans : **(Imp.)**

Following points highlights the application of Baye's theorem,

1. In Baye's theorem, posterior probabilities can be known by rivising priori probabilities with the help of new information
2. The probability of occurrence of future events can be known by Baye's theorem.
3. Baye's theorem offers a powerful statistical tool.
4. Baye's theorem helps the business and management executives to take effective decisions in uncertain situation.

Baye's theorem is also known as 'Probability of Causes' as it helps in determining the probability which a particular effect has due to a specific cause.

PROBLEMS

28. In a certain college, 25% of boys and 10% of girls are studying mathematics. The girls constitute 60% of the student body,

(a) What is the probability that mathematics is being studied ?

(b) If a student is selected at random and is found to be studying mathematics, find the probability that the student is a girl?

(c) Probability of maths student is a boy

Sol :

Given P (Boy) = P(B) = $\frac{40}{100} = \frac{2}{5}$

and P (Girl) = P(G) = $\frac{60}{100} = \frac{3}{5}$

Probability that mathematics is studied given that the student is a boy = P{M / B) = $\frac{25}{100} = \frac{1}{4}$

Probability that mathematics is studied given that the student is a girl = P(M / G) = $\frac{10}{100} = \frac{1}{10}$

(a) Probability that the student studied Mathematics = P(M)

= P(G) P(M/G) + P(B) P(M/B)

∴ By total probability theorem,

$$P(M) = \frac{3}{5}.\frac{1}{10} + \frac{2}{5}.\frac{1}{4}$$

$$= \frac{4}{25}$$

(b) By Baye's theorem, probability of mathematics student is a girl = P(G / M)

$$= \frac{P(G)P(M/G)}{P(M)} = \frac{\frac{3}{5}.\frac{1}{10}}{\frac{4}{25}} = \frac{3}{8}$$

(c) Probability of maths student is a boy = P(B / M)

$$= \frac{P(B)P(M/B)}{P(M)} = \frac{\frac{2}{5}.\frac{1}{4}}{\frac{4}{25}} = \frac{5}{8}$$

29. A bag A contains 2 white and 3 red balls and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the red ball drawn is from bag.

Sol :

Let A and B denote the events of selecting bag A and bag B respectively.

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

Let R denote the event of drawing a red ball.

Having selected bag A, the probability to draw a red ball from A = P(R/ A) = $\frac{3}{5}$

Similarly $P(R/A) = \frac{5}{9}$

One of the bags is selected at random and from it a ball is drawn at random.

It is found to be red. Then the probability that the selected bag is B

$$= P(B).P(R/B) = \frac{P(B).P(R/B)}{P(A).P(R/A) + P(B)P(R/B)}$$

$$= \frac{\frac{1}{2}.\frac{5}{9}}{\frac{1}{2}.\frac{3}{5} + \frac{1}{2}.\frac{5}{9}} = \frac{25}{52}$$

30. A factory has two machines. Empirical evidence has established that machines I and II produce 30% and 70% of output respectively. It has also been established that 5% and 1% of the output produced by these machines respectively was defective. A defective item is drawn at random. What is the probability that the defective item was produced by either machine I or machine II ?

Sol : **(Jan.-21, Imp.)**

Let A_1 be the event of drawing of an item produced by machine1.

A_2 is the event of drawing an item produced by machine 2

$P(A_1)$ is the probability of getting an item produced by machine 1

$P(A_2)$ is the probability of getting an item produced by machine 2.

$P(B/A_1)$ is the probability of getting defective machine 1

$P(B/A_2)$ is the probability of getting defective machine 2.

From the given data

$P(A_1) = 0.3$ (30%) $P(B/A_1) = \frac{5}{100} = 0.05$

$P(Q_2) = 0.7$ (70%) $P(B/A_2) = 1\% = \frac{1}{100} = 0.01$

Computation of posterior probabilities

Events	$P(A_1)$	$P(B/A_1)$	$P(A \cap B)$	$P(A_1/B)$
A_1	$P(A_1) = 0.3$	$P(B/A_1) = 0.05$	$0.3 \times 0.05 = 0.015$	$\frac{0.015}{0.022} = 0.682$
A_2	$P(A_2) = 0.7$	$P(B/A_2) = 0.01$	$0.7 \times 0.01 = 0.0007$	$\frac{0.007}{0.022} = 0.318$
	1.00	0.06	0.022	1.000

31. In a bolt factory, the Machines P,Q and R manufacture respectively 25%, 35% and 40% of the total of their outputs 5,4,2 percents respectively are defective bolts. A bolt is drawn at random from the product, and is known to be defective, What are the probabilities that it was manufactured by the machines P,Q and R.

Sol : **(June-18)**

Let P(A), P(B), P(C) be the probabilities of the Events that the bolts are manufactured by the machines A, B & C respectively. Then

$$P(A) = \frac{25}{100} = 0.25, \quad P(B) = \frac{35}{100} = 0.35, \quad P(C) = \frac{40}{100} = 0.40$$

Let 'D' denotes that the bolts is defective. Then

$$P(D/A) = \frac{5}{100} = 0.05, \quad P(D/B) = \frac{4}{100}, \quad P(D/C) = \frac{2}{100}$$

(i) If bolt is defective, then the probability that it is from machine A.

By using Baye's Theorem

$$P(A/D) = \frac{P(D/A)P(A)}{P(D/A)P(A)+P(D/B)P(B)+P(D/C)P(C)}$$

$$= \frac{0.05(0.25)}{0.05(0.25)+0.04(0.35)+0.02(0.4)}$$

$$P(A/D) = \frac{0.0125}{0.0125+0.014+0.008} = \frac{0.0125}{0.0345} = 0.362$$

(ii) If bolt is defective, then the probability that it is from mechine B.

$$P(B/D) = \frac{P(D/B)P(B)}{P(D/A)P(A)+P(D/B)P(B)+P(D/C)P(C)}$$

$$= \frac{0.04(0.35)}{0.05(0.25)+0.04(0.35)+0.02(0.40)}$$

$$P(B/D) = \frac{0.014}{0.0345} = 0.4057$$

(iii) It bolt is defective, then the probability that it is from machine C

$$P(C/D) = \frac{P(C/D)P(C)}{P(D/A)P(A)+P(D/B)P(B)+P(D/C)P(C)}$$

$$\frac{(0.02)(0.40)}{0.05(0.25)+0.04(0.35)+0.02(0.40)} = \frac{0.008}{0.0345}$$

$$P(C/D) = 0.2318.$$

32. A company has two plants for manufacturing scooters. Plant I manufacturers 80% of the Scooters and Plant II manufactures 20%. At the Plant I 85% Scooters are rated to be of standard quality and at plant II 65% Scooters are rated to be of standard quality. One Scooter was selected at random. What is the probability that

i) It is manufactured by Plant I

ii) It is manufactured by Plant II – which is of standard quality.

Sol : **(June-19, Imp.)**

Let $P(E_1)$ denoted as manufacturing scooters from plant - I.

Probability of plant-I manufacturing scooters = 80% = 0.80.

$$\therefore P(A) = 0.8$$

Let rated $P\left(\frac{A}{E_1}\right)$ denotd as scooters are rated as standard quality.

Probability of scooters rated as standard quality in plant I = 85% = 0.85.

$$P\left(\frac{A}{E_1}\right) = 0.85$$

Let $P(E_2)$ denoted as manufacturing scooters from plant - II.

Probability of plant - II manufacturing scooters = 20% = 0.20.

$$\therefore P(E_2) = 0.20.$$

Let $P\left(\frac{A}{E_2}\right)$ denoted as scooters are rated standard quality in plant II.

Probability of scooters rated as standard quality in plant II is 65% = 0.65.

$$P\left(\frac{A}{E_2}\right) = 0.65$$

(i) **It is manufactured by plant I**

$$P\left(\frac{E_1}{A}\right) = \frac{P(E_1)\,P\left(\frac{A}{E_1}\right)}{P(E_1)\,.\,P\left(\frac{A}{E_1}\right) + P(E_2).P\left(\frac{A}{E_2}\right)}$$

$$= \frac{0.80 \times 0.85}{(0.80 \times 0.85) + (0.2) \times (0.65)}$$

$$= \frac{0.68}{0.68 + 0.13}$$

$$= \frac{0.68}{0.81}$$

$$\therefore \quad P\left(\frac{E_1}{A}\right) = 0.839.$$

(ii) **It is manufactured by Plant - II**

$$P\left(\frac{E_2}{A}\right) = \frac{P(E_2).\,P\left(\frac{P}{E_2}\right)}{P(E_1).\,P\left(\frac{A}{E_1}\right) + P(E_2).\,P\left(\frac{A}{E_2}\right)}$$

$$= \frac{(0.20)\times(0.65)}{(0.80\times 0.85)+(0.2)(0.65)}$$

$$= \frac{0.13}{0.68+0.13} = \frac{0.13}{0.81} = 0.1604.$$

Short Question and Answers

1. Dependent Event.

Ans :

Dependent events are those in which the occurrence or non-occurrence of one event in any one trial affects the probability of other events in other trials. For example, if a card is drawn from a pack of playing cards and is not replaced, this will alter the probability that the second card drawn is, say an ace. Similarly, the probability of drawing a queen from a pack of 52 cards is $\frac{4}{52}$ or $\frac{1}{13}$. But if the card drawn (queen) is not replaced in the pack, the probability of drawing again a queen is 3/51 (\ the pack now contains only cards out of which there are 3 queens).

2. Independent Event

Ans :

Two or more events are said to be independent when the outcome of one does not affect, and is not affected by the other.

For example, if a coin is tossed twice, the result of the second throw would in no way be affected by the result of the first throw. Similarly, the results obtained by throwing a dice are independent of the results obtained by drawing an ace from a pack of cards.

To consider two events that are not independent, let A stand for a firm's spending a large amount of money on advertisement and *B* for its showing an increase in sales. Of course, advertising does not guarantee higher sales, but the probability that the firm will show an increase in sales will be higher if A has taken place.

3. Explain the Axiomatic approach to probability.

Ans :

The axiomatic approach to probability was introduced by the Russian mathematician A. N. Kolmogorov in the year 1933. Kolmogorov axiomised the theory of probability and his book Foundations of Probability, published in 1933, introduces probability as a set function and is considered as a classic. When this approach is followed, no precise definition of probability is given, rather we give certain axioms or postulates on which probability calculations are based. The whole field of probability theory for finite sample spaces* is based upon the following three axioms :

1. The probability of an event ranges from zero to one. If the event cannot take place its probability shall be zero and if it is certain, i.e., bound to occur, its probability shall be one.
2. The probability of the entire sample space is 1, Le., P(S) = 1.
3. If A and B are mutually exclusive (or disjoint) events then the probability of occurrence of either A or B denoted by $P(A \cup B)$ shall be given by :

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

   It may be pointed out that out of the four interpretations of the concept of probability, each has its own merits and one may use whichever approach is convenient and appropriate for the problem under consideration.

The probability of a event A, denoted by P(A) is so chosen as to satisfy the following three axioms.

i) $P(A) \geq 0 \Rightarrow$ This axiom states that the probability of occurrence of an event A in a random experiment may be zero or any positive number and it must not be negative number.

ii) P(S) - 1 $\Rightarrow$ This states that the sample space, S, itself is an event and since it is the event comprising all possible outcomes, it should have the highest possible probability, i.e., one.

iii) If $A \cap B = 0$, Then $P(A \cup B) = P(A) + P(B) \Rightarrow$ This axiom states that the probability of the event equal to the union of any number of mutually exclusive events is equal to the sum of the individual even probabilities.

4. Mutually Exclusive Events

Ans :

Two events are said to be mutually exclusive or incompatible when both cannot happen simultaneously in a single trial or, in other words, the occurrence of any one of them precludes the occurrence of the other.

For example, if a single coin is tossed either head can be up or tail can be up, both cannot be up at the same time. Similarly, a person may be either alive or dead at a point of time he cannot be both alive as well as dead at the same time.

To take another example, if we toss a dice and observe 3, we cannot expect 5 also in the same toss of dice. Symbolically, if A and B are mutually exclusive events, P (AB) = 0.

5. Non-mutually Exclusive Events.

Ans :

When two events can occur simultaneously in a single trial then such events are said to be non-mutually exclusive events.

Eaxample, from a pack of cards, drawing a red card and drawing a queen are the two events. These two events can occur simultaneously while drawing a red queen.

Hence, these two events are said to be non-mutually exclusive events which can occur at the same time.

6. What is probability?

Ans :

Introduction

An Italian mathematician, Galileo (1564 - 1642), attempted a quantitative measure of probalmhjfwhile dealing with some problems related to gambling. In the middle of 17th Century, two French mathematicians, Pascal and Fermat, laid down the first foundation of the mathematical theory of probability while solving the famous 'Problem of Points' posed by Chevalier-De-Mere. Other mathematicians from several countries also contributed in no small measure to the theory of probability. Outstanding of them were two Russian mathematicians, A. Kintchine and A. Kolmogoroff, who axiomised the calculus of probability.

If an experiment is repeated under similar and homogeneous conditions, we generally come across two types of situations.

(i) The net result, what is generally known as 'outcome' is unique or certain.

(ii) The net result is not unique but may be one of the several possible outcomes.

The situations covered by :

(i) are known as 'deterministic' or 'predectable' and situations covered by

(ii) are known as 'probabilistic' or 'unpredictable'.

'Deterministic' means the result can be predicted with certainty. For example, if r is the radius of the sphere then its volume is given by $V = \frac{4}{3}\pi r^3$ which gives uniquely the volume of the sphere.

There are some situations which do not lend themselves to the deterministic approach and they are known as 'Probabilistic'.

For example, by looking at the sky, one is not sure whether the rain comes or not.

In such cases we talk of chances or probability which can be taken as a quantitative measure of certainity.

Definitions

In a random experiment, let there be n mutually exclusive and equally likely elementary events. Let E be an event of the experiment. If m elementary events form event E (are favourable to E), then the probability of E (Probability of happening of E or chance of E), is defined as

$$P(E) = \frac{m}{n} = \frac{\text{Number of elementary events in E}}{\text{Total number of elementary events in the random experiment}}$$

If $\bar{E}$ denotes the event of non-occurrence of E, then the number of elementary events in E is n-m and hence the probability of E (non-occurrence of E) is

$$P(E) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(E) \Rightarrow P(E) + P(\bar{E}) = 1$$

Since m is a non-negative integer, n is a positive integer and $m \leq n$, we have m

$$0 \leq \frac{m}{n} \leq 1.$$

Hence $0 \leq P(E) \leq 1$ and $0 \leq P(\bar{E}) \leq 1$.

7. Random Experiment.

Ans :

If an 'experiment' is conducted, any number of times, under essentially identrfcal conditions, there is a set of all possible outcomes associated with it. If the result is not certain and is anyone of the several possible outcomes, the experiment is called a random trial or a random experiment. The outcomes are known as elementary events and a set of outcomes is an event. Thus an elementary event is also an event.

8. Explain the importance of probability.

Ans :

(i) The probability theory is very much helpful for making prediction. Estimates and predictions form an important part of research investigation. With the help of statistical methods, we make estimates for the further analysis. Thus, statistical methods are largely dependent on the theory of probability.

(ii) It has also immense importance in decision making.

(iii) It is concerned with the planning and controlling and with the occurrence of accidents of all kinds.

(iv) It is one of the inseparable tools for all types of formal studies that involve uncertainty.

(v) The concept of probability is not only applied in business and commercial lines, rather than it is also applied to all scientific investigation and everyday life.

(vi) Before knowing statistical decision procedures one must have to know about the theory of probability.

(vii) The characteristics of the Normal Probability. Curve is based upon the theory of probability.

9. Define set.

Ans :

A set is a well-defined collection of all possible objects according to a well defined rule. The objects comprising a set are called elements or members of the set. A subset, B of a set A is another set whose element are also elements of the set A and is written as $B \subset A$. The set B is a proper subset of A if B is a subset of A and at least one element of A is not contained in B.

Two sets A and B are said to be equal or identical if $A \subset B$ and $B \subset A$, i.e., $A = B$ or $B = A$.

If a set is having no elements at all, it is said to be an empty set or null set and is usually denoted by ϕ. Thus, empty set is a subset of every set. The universal set, S is the set of all elements considered in a given problem.

If there are three sets A, B and C in such a way that $A \subset B$ and $B \subset C$, then $A \subset C$. This property is called transitivity.

A set is said to be finite if it is empty or contains n elements $(n > 0)$ otherwise it is infinite.

10. Permutation

Ans :

Arrangement of 'n' things in a specified order is called permutation. Here all things are taken at a time.

Example

Consider the letters a, b and c. Considering all the three letters at a time, the possible permutations are abc, acb, bca, bac, cba and cab.

Arrangement or ‘r’ things taken at a time from ‘n’ things where $r \leq n$, in a specified order is called r-permutation.

Example : From the three letters a, b and c, the possible 2-permutations i.e., 2 letters taken at a time are ab, ba, ac, ca, bc, cb.

The above computation is direct computation. Now we apply the counting technique. Consider the permutations considering all the three letters at a time.

The first letter can be selected in three different ways. Following this, the second letter can be selected in two different ways. And the 3rd letter can be selected in only one way. Let the three letter word be represented as,

c	a	b

The possible ways of selecting the letters is

3	2	1

Thus, the possible No. of permutations are 3.2.1 = 6. Now, consider 2-permutation.

3	2

∴ The possible No. of permutations are 3.2 = 6

The number of permutations taking ‘r’ things at a time from ‘n’ available things is denoted as p(n, r) or ${}^{n}P_{r}$.

11. Combination

Ans :

In permutations, the order of arrangement of objects is important. But, in combinations, order is not important, but only selection of objects.

For example, abc, bca, cab are different permutations but only one combination of 3 letters a, b and c.

The possible number of combinations of n objects, taken r at a time is denoted by C(n, r) or n_c.

Consider 3 letter a, b and c. The possible combinations taking two at a time are ab, bc, ac, ca, bc, cb. Thus, the number of permutations is equal to the number of combinations multiplied by 2!

$$\therefore \quad P(n, r) = {}^{n}P_{r} = r!\,{}^{n}C_{r} = r!\,{}^{n}C_{r} \Rightarrow {}^{n}C_{r} = \frac{P(n,r)}{r!} = \frac{\angle n}{\angle r.\angle n-r}$$

Ordered Partitions

Assume that there are 10 toys to be divided between 4 children such that one of them gets 4 and remaining each gets 2 toys.

So, the given set A(=10) should be divided into ordered partions $A_1(=4)$, $A_2(=2)$ and (= 2).

$$\{A\} = \{A_1, A_2, A_3, A_4\}$$

From 10 toys, selecting the first 4 is in ${}^{10}C_4$ ways, i.e., A_1 can be determined in ${}^{10}C_4$ ways. Following this, from the remaining 6, A_2 can be determined in ${}^{6}C_2$ ways. Following this, A_3 can be determined in ${}^{4}C_2$ ways. Following this A_4 can be determined in ${}^{2}C_2$ ways.

Thus, there are ${}^{10}C_4 . {}^{6}C_4 . {}^{4}C_2 . {}^{2}C_2$ different ordered partitions of A into A_1 consisting of 4 toys, A_2 consisting of 2 toys, A_3 consisting of 2 toys and A_4 consisting of 2 toys.

$${}^{10}C_4 . {}^{6}C_2 . {}^{4}C_2 . {}^{2}C_3 = \frac{10!}{4!6!} \times \frac{6!}{2!4!} \times \frac{4!}{2!2!} \times \frac{2!}{2!0!} \times \frac{10!}{4!2!2!2!} = 18900$$

Thus, this can be generalized as $\frac{n!}{n_1!n_2!...n_K!}$, where

A_1 contains n_1 elements, A_2 contains n_2 elements and so on.

Exercise Problems

1. Two aeroplanes bomb a target in succession. The probability of each correctly scoring a hit is 0.3 and 0.2 respectively. The second will bomb only if the first misses the target. Find the probability that (i) target is hit (ii) both fails to score hits.

   [Ans : (i) 0.44, (ii) 0.56]

2. Determine (i) $P\left(\frac{B}{A}\right)$ (ii) $P\left(\frac{A}{B^C}\right)$ if A and B are events with $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{4}$, $P(A \cup B) = \frac{1}{2}$.

   [Ans : $\frac{1}{2}$]

3. Box A contains 5 red and 3 white marbles and box B contains 2 red and 6 white marbles. If a marble is drawn from each box, what is the probability that they are both of same colour.

   [Ans : 0.109]

4. Two marbles are drawn in succession from a box containing 10 red, 30 white, 20 blue and 15 orange marbles, with replacement being made after each draw. Find the probability that

   (i) Both are white

   (ii) First is red and second is white

   [Ans : (i) $\frac{4}{25}$, (ii) $\frac{4}{75}$]

5. In a factory, machine A produces 40% of the output and machine B produces 60%. On the average, 9 items in 1000 produced by A are defective and 1 item in 250 produced by B is defective. An item drawn at random from a day's output is defective. What is the probability that it was produced by A or B ?

   [Ans : 0.4]

6. A manufacturer firm produces steel pipes in three plants, with daily production volume of 500; 1000 and 2000 units respectively. According to past experience; it is known that the fraction of defective outputs produced by these plants are respectively 0.005; 0.008 and 0.010. If a pipe is selected from day's total production and found to be defective, find out (i) from which plant the pipe came (ii) what is the probability that it came from the first plant ?

   [Ans : $\frac{40}{61}$]

Choose the Correct Answer

1. Way of getting information from measuring observation whose outcomes occurrence is on chance is called [b]

   (a) Beta experiment (b) Random experiment

   (c) Alpha experiment (d) Gamma experiment

2. Probability of second event in situation if first event has been occurred is classified as [b]

   (a) Series probability (b) Conditional probability

   (c) Joint probability (d) Dependent probability

3. Probability which is based on self-beliefs of persons involved in experiment is classified as [a]

   (a) Subjective approach (b) Objective approach

   (c) Intuitive approach (d) Sample approach

4. In probability theories, events which can never occur together are classified as [c]

   (a) Collectively exclusive events (b) Mutually exhaustive events

   (c) Mutually exclusive events (d) Collectively exhaustive events

5. Number of individuals arriving at boarding counter on an airport is an example of [a]

   (a) Numerical outcome (b) Non numerical outcome

   (c) Random outcome (d) Simple outcome

6. If two events X and Y are considered as partially overlapping events then rule of addition can be written as [d]

   (a) P(X or Y) = P(X) - P(Y) + P(X and Y)

   (b) P(X or Y) = P(X) + P(Y) * P(X - Y)

   (c) P(X or Y) = P(X) * P(Y) + P(X - Y)

   (d) P(X or Y) = P(X) + P(Y) - P(X and Y)

7. According to combination rule, if total number of outcomes are ‘r’ and distinct outcome collection is ‘n’ then combinations are calculated as [a]

   (a) n! D r!(n - r)! (b) n! D r!(n + r)!

   (c) r! D n!(n - r)! (d) r! D n!(n + r)!

8. For a random experiment, all possible outcomes are called [d]

(a) Numerical space (b) Event space

(c) Sample space (d) Both b and c

9. Types of probabilities for independent events must includes [d]

(a) Joint events (b) Marginal events

(c) Conditional events (d) All of above

10. Probability without any conditions of occurrence of an event is considered as [b]

(a) Conditional probability (b) Marginal probability

(c) Non conditional probability (d) Occurrence probability

Fill in the blanks

1. Marginal probability of independent events and dependent events must be ______.
2. Method of counting outcomes in which number of outcomes are determined while considering ordering is classified as ______.
3. In probability theory, events are denoted by ______.
4. Difference between sample space and subset of sample space is considered as ______.
5. Occurrence of two events in a way that events have some connection in between is classified as ______.
6. Method in which previously calculated probabilities are revised with new probabilities is classified as ______.
7. Probability of events must lie in limits of ______ .
8. Measure of chance of an uncertain event in form of numerical figures is classified as ______.
9. Events in which some points of sample are common are considered as ______.
10. In a box, there are 8 red, 7 blue and 6 green balls. One ball is picked up randomly. What is the probability that it is neither red nor green is ______.
11. Let S be a sample space and two mutually exclusive events A and B such that $A \cup B = S$. If P(.) denotes the probability of the events, the maximum value of P(A) P(B) is ______.
12. Three fair cubical dice are thrown simultaneously The probability that all three dice have the same number of dots on the faces showing up is (upto third decimal place) ______.
13. Two coins are tossed simultaneously. The probability (upto two decimal points accuracy) of getting at least one head is ______.

ANSWERS

1. Same
2. Permutation
3. Capital letters
4. Complementary events
5. Compound Events
6. Bayes theorem
7. Zero to One
8. Probability
9. Overlapping Events
10. 1/3
11. 0.25
12. 0.028
13. 0.75

UNIT V	**THEORITECAL DISTRIBUTIONS:** Binomial Distribution: Importance - Conditions - Constants - Fitting of Binomial Distribution. Poisson Distribution: - Importance - Conditions - Constants - Fitting of Poisson Distribution. Normal Distribution: - Importance - Central Limit Theorem - Characteristics - Fitting a Normal Distribution (Areas Method Only).

5.1 Binomial Distribution

5.1.1 Importance

Q1. What is binomial distribution? State the importance of binomial distribution.

Ans : **(Imp.)**

Meaning

Binomial distribution was discovered by James Bernoulli in the year 1700 and it is a discrete probability distribution.

The Binomial distribution can be described with the help of a function and is derived as follows:

Let the number of trials be n. The trials be independent i.e., the success or failure at one trial does not affect the outcome of the other trials. Thus the probability of success remains the same from trial to trial. Also let 'p' be the probability of success and 'q' be the probability of failure. Then we have $p + q = 1$ or $q = 1 - p$.

One could be interested in finding the probability of getting r successes, which implies getting (n – r) failures.

One of the ways in which one can get r successes and (n – r) failure is to get r continuous successes first and (n – r) successive failures.

The probability of getting such a sequence of r successes and (n-r) failures is, $= p^r q^{n-r}$ (using Multiplication Theorem of Probability)

However, the number of ways in which one can get r successes and n – r failures is ^nC_r and the probability for each of these ways is $p^r q^{n-r}$.

Thus, if x is the random variable representing the number of successes, the probability of getting r successes and n – r failures, in n trials, is given by the probability function

$$P(x = r) = {}^nC_r\, p^r q^{n-r},\ r = 0, 1, 2, ..., n$$

This probability function is popularly known as the Binomial Distribution.

The probabilities of 0, 1, 2, ..., r, ..., n successes are therefore given by

$$q^n,\ {}^nC_1\, p\, q^{n-1},\ {}^nC_2\, p^2 q^{n-2},\ ...,\ {}^nC_r\, p^r q^{n-r},\ ...,\ p^n.$$

The probability of the number of successes so obtained is called the Binomial Probability Distribution, because these probabilities are the successive terms in the expansion of the binomial $(q + p)^n$. This distribution contains two independent constants namely n and p (or q). They are called parameters of the Binomial Distribution. Sometimes, n is also known as the degree of the distribution.

Examples

(i) The number of defective bolts in a box containing bolts.

(ii) The number of machines lying idle in a factory having machines.

(iii) The number of post-graduates in a group of men.

(iv) The number of oil wells yielding natural gas in a group of wells test drilled.

Importance

The binomial probability distribution is a discrete probability distribution that is useful in describing an enormous variety of real life events.

For example, a quality control inspector wants to know the probability of defective light bulbs in a random sample of 10 bulbs if 10 per cent of the bulbs are defective. He can quickly obtain the answer from tables of the binomial probability distribution. The binomial distribution can be used when :

1. The outcome or results of each trial in the process are characterized as one of two types of possible outcomes. In other words, they are attributes.
2. The possibility of outcome of any trial does not change and is' independent of the results of previous trials.

Q2. What are the properties of binomial distribution.

Ans :

The properties of Binomial distribution are as follows,

1. It describes the distribution of probabilities when there are only two mutually exclusive outcomes for each trial of an experiment for example while tossing a coin, the two possible outcomes are head and tail.
2. The process is performed under identical conditions for 'n' number of times.
3. Each trial is independent of other trials. It means the outcome of a particular trial does not affect the outcome of another trial.
4. The probability of success 'p' remains same for trial to trial throughout the experiment and similarly, the probability of failure (q = 1 –p) also remains constant overall the observations.

5.1.2 Conditions

Q3. What are the conditions of binomial distribution?

Ans :

Binomial Distribution holds under the following conditions :

1. Trials are repeated under identical conditions for a fixed number of times, say *n* times.
2. There are only two possible outcomes, *e.g.* success or failure for each trial.
3. The probability of success in each trial remains constant and does not change from trial to trial.
4. The trials are independent *i.e.,* the probability of an event in any trial is not affected by the results of any other trial.

5.1.3 Constants

Q4. What are the constants of binomial distribution?

Ans :

The random variable 'x' the number of successes 'r' in n trials has a probability distribution.

$$P(x = r) = {}^nC_r\, p^r\, q^{n-r}$$

Where,

r = 1, 2, ... n

p = Probability of successes on a single trail

q = Probability of failure on a single trial

n = Number of Bernoulli trials

Following are the constants of binomial distribution,

(i) For binomial distribution variance is less than mean.

(ii) For binomial distribution with parameters n and p variance cannot exceed n/4.

(iii) Mean and variance of a binomial random variable depend on values assumed by parameters n, p and q.

the various constants of the binomial distribution can be listed in the following table,

Mean	=	np
Standard Deviation	=	$\sqrt{npq}$
Variance (μ_2)	=	npq
Skewness (μ_1)	=	$\frac{q-p}{\sqrt{npq}}$
Kurtosis (μ_4)	=	$\frac{1-6pq}{npq}$

5.1.4 Fitting of Binomial Distribution

Q5. Briefly describe about Fitting a Binomial Distribution.

(OR)

Explain about Fitting a Binomial distribution.

Ans :

When a binomial distribution is to be fitted to observe data, the following procedure is adopted:

1. Determine the values of p and q. If one of these values is known the other can be found out by the simple relationship p = (1 – q). and q = (1– p). When p and p are equal the distribution is symmetrical, for p and q may be interchanged without alternating the value of any terms, and consequently terms equidistant from the two ends of the series are equal. If p and q are unequal, the distribution is skew. If p is less than $\frac{1}{2}$, the distribution is positively skewed and when p is more than $\frac{1}{2}$ the distribution is negatively skewed.
2. Expand the binomial $(q + p)^n$. The power n is equal to one less than the number of terms in the expanded binomial. Thus when two coins are tossed (n = 2) there will be three terms in the binomial. Similarly, when four coins are tossed (n = 4) there will be five terms, and so on.
3. Multiply each term of the expanded bionominal by N (the total frequency), in order to obtain the expected frequency in each category.

PROBLEMS

1. A fair coin is tossed six times. Find the probability of getting four heads.

Sol :

p = probability of getting a head = 1/2

q = probability of not getting head = 1/2

and n = 6, r = 4

We know that $P(r) = {}^nC_r\, p^r\, q^{n-r}$

$$\therefore \quad P(4) = {}^6C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{6-4}$$

$$= \frac{6!}{4!2!}\cdot\left(\frac{1}{2}\right)^6$$

$$= \frac{6\times5}{2}\cdot\frac{1}{2^6} = \frac{15}{64}$$

= 0.2344.

2. Determine the probability of getting the sum 6 exactly 3 times in to throws with a pari of fair dice.

Sol :

In a single throw of a pair of fair dice, a sum of 6 can occur in 5 ways: (1, 5), (5, 1), (2, 4), (4, 2) and (3, 3) out of 6 × 6 = 36 ways. Thus,

p = Probability of occurrence of 6 in one throw = $\frac{5}{36}$

$q = 1 - p = 1 - \frac{5}{36} = \frac{31}{36}$

n = Number of trials = 7

∴ Probability of getting 6 exactly thrice in 7 throws

$$= {}^7C_3\, p^3\, q^{7-3} = {}^7C_3\left(\frac{5}{36}\right)^3\left(\frac{31}{36}\right)^4$$

$$= \frac{35(125)\,(31)^4}{(36)^7}$$

= 0.0516 (nearly).

3. If the probability of a defective bolt is 1/8, find :

(i) The Mean

(ii) The variance for the distribution of defective bolts of 640.

Sol :

We are given

p = The probability of a defective bolt = 1/8 and n = 640.

∴ Mean of the distribution,

$$\mu = np = \frac{640}{8} = 80$$

Also $q = 1 - p = 1 - \frac{1}{8} = \frac{7}{8}$

Hence variance of the distribution = npq = (np)q = $\mu q = 80 \times \frac{7}{8} = 70$.

4. 6 coins are tossed at a time what is the probability of obtaining 4 or more Heads.

Sol: **(June-19)**

The Distribution is a binomial distribution

Here n = 6

Let the probability of getting a head in one toss = P

Since the coin is fair P = 1/2

i.e., change of success - P

chance of failure = q = 1 – P

The probability of getting x number of success in 'n' number of tosses is

$$P(x = x) = nc_x p^x q^{n-x}$$

4 or more Heads $P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6)$

$$= {}^6C_4 (1/2)^4 (1/2)^{6-4} + {}^6C_5 (1/2)^5 (1/2)^{6-5} + {}^6C_6 (1/2)^6 (1/2)^{6-6}$$

15(0.0625) (0.25) + 6 (0.03125) (0.5) + 1(0.015625) (1)

0.2343 + 0.09375 + 0.01562

$$\boxed{P(x \geq 4) = 0.34325}$$

∴ The probability of obtaining 4 or more heads is 0.34325

5. Comment on the following:

For a Binomial Distribution Mean = 7 and Variance = 11.

Sol : **(Jan21, June-19)**

Given mean (np) = 7

Variance (npq) = 11

npq = 11

7q = 11

$$q = \frac{11}{7} = 1.57$$

The given statement is not correct because the value of q is more than 1, it should not exceed 1.

6. If the probability of a defective bot is 0.2, find (i) mean (ii) standard deviation for the distribution of bolts in a total of 400.

Sol :

Given n = 400, p = 0.2. $\therefore$ q = 1 – p = 1 – 0.1 = 0.8

(i) Mean = np = 400 (0.2) = 80 ... (1)

(ii) S.D. = $\sqrt{npq} = \sqrt{80(0.8)} = \sqrt{64}$ = 8 [by (1)]

7. A die is tossed thrice. A success is getting 1 or 6 on a toss. Find the mean and variance of the number of successes.

Sol : **(June-19, Imp.)**

The probability of success, $p = \frac{2}{6} = \frac{1}{3}$

$\therefore$ The probability of failure, $q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

No. of trials, n = 3

$\text{Mean} = np = 3\left(\frac{1}{3}\right) = 1$

$\text{Variance} = npq = (np)q = (1)\, q = q = \frac{2}{3}$

8. 8 Coins are tossed at a time, 256 times. Find the expected frequencies of successes Getting a Head) and tabulate the results obtained.

Sol : **(Jan.-21, Imp.)**

n = 8

N = 256

The probability of getting a head (p) = $\frac{1}{2}$

The probability of getting a tail (q) = $\frac{1}{2}$

$= n_{C_r}\, q^{n-r}\, p^r$

$= 8C_r \left(\frac{1}{2}\right)^{8-r} \left(\frac{1}{2}\right)^{r} = 8C_r \left(\frac{1}{2}\right)^{r}$

The frequencies of 0, 1, 2, 3....8 are as follows

Success	N × p(r)	Expected Frequency
0	$256\left(\frac{1}{256}\times 8C_0\right)$	1
1	$256\left(\frac{1}{256}\times 8C_1\right)$	8
2	$256\left(\frac{1}{256}\times 8C_2\right)$	28
3	$256\left(\frac{1}{256}\times 8C_3\right)$	56
4	$256\left(\frac{1}{256}\times 8C_4\right)$	70
5	$256\left(\frac{1}{256}\times 8C_5\right)$	56
6	$256\left(\frac{1}{256}\times 8C_6\right)$	28
7	$256\left(\frac{1}{256}\times 8C_7\right)$	8
8	$256\left(\frac{1}{256}\times 8C_8\right)$	1
		256

9. Five coins are tossed 3,200 times, find the frequencies of the distribution of heads and tails, and tabulate the results and also calculate Mean and standard Deviation of fitted distribution.

Sol: **(June-18)**

The coin is unbiased

$\therefore\ p=\frac{1}{2}, q=\frac{1}{2}$ and $n=5$

$N = 3200$

No. of heads x	P(x) Probability $^nC_x p^x q^{n-x}$	Expected Frequency f(x) = N * P(x)
0	$^5C_0(1/2)^o(1/2)^{5-0} = 0.03125$	$3200(0.03125) = 100$
1	$^5C_1(1/2)^1(1/2)^{5-1} = 0.15625$	$3200(0.15625) = 500$
2	$^5C_2(1/2)^2(1/2)^{5-2} = 0.3125$	$3200(0.3125) = 1000$
3	$^5C_3(1/2)^3(1/2)^{5-3} = 0.3125$	$3200(0.3125) = 1000$
4	$^5C_4(1/2)^4(1/2)^{5-4} = 0.15625$	$3200(0.15625) = 500$
5	$^5C_5(1/2)^5(1/2)^{5-5} = \frac{0.03125}{1}$	$3200(0.03125) = \frac{100}{3200}$

Given frequency-Expected frequency.

i) Mean of Binomial Distribution $(\mu) = np$

$$5(0.5)$$

$$\boxed{\mu = 2.5}$$

ii) Standard Deviation of Binomial Distribution $(\sigma) = \sqrt{npq}$

$$\sqrt{5(0.5)(0.5)}$$

$$\boxed{\sigma = 1.118}$$

10. Ten unbiased coins are tossed simultaneously. Find the probability of obtaining:

i) Exactly 6 Heads

ii) Atleast 8 Heads

iii) No Heads

iv) Atleast one Head

v) Not more than 3 Heads and

vi) Atleast 4 heads.

Sol : **(June-19, Imp.)**

Given n = 10, p = 0.5, q = 0.5

(i) Exactly 6 heads

$\therefore$ r = 6

$$p(x = r) = n_{c_r} \cdot p^r \cdot q^{n-r} \left[n_{c_r} = \frac{n!}{r!(n-r)!} \right]$$

$p(x = 6) = 10_{c_6} . (0.5)^6 . (0.5)^{10-6}$

$= \frac{10 \times 9 \times 8 \times 7 \times \cancel{6 \times .... \times 1}}{\cancel{6 \times .... \times 1} \times 4 \times 3 \times 2 \times 1} \times 0.015625 \times 0.0625$

$= \frac{5040}{24} \times 0.015625 \times 0.0625$

$= 210 \times 0.015625 \times 0.0625 = 0.2050.$

(ii) At least 8 head

$r \geq 8$ (r = 8 + r = 9 + r = 10)

$p(r = 8) = 10_{c_8} . (0.5)^8 . (0.5)^{10-8}$

$= \frac{10 \times 9 \times \cancel{8 \times ..... \times 1}}{\cancel{8 \times ....} \times 1 \times 2 \times 1} \times 0.00390625 \times 0.25$

$= \frac{90}{2} \times 0.00390625 \times 0.25$

$= 45 \times 0.00390625 \times 0.25 = 0.044578125$

r = 9

$p(r = 9) = 10_{C_9} . (0.5)^9 \times (0.5)^{10-9}$

$= \frac{10 \times \cancel{9 \times 8 \times 7.... \times 1}}{\cancel{9 \times ... = - \times 1} \times 1} \times 0.001953125 \times (0.5)^1.$

$= 10 \times 0.001953125 \times 0.5 = 0.009765625$

If r = 10

$p(r = 10) = 10_{C_{10}} \times (0.5)^{10} \times (0.5)^{10-10}$

$= \frac{\cancel{10 \times ...... \times 1}}{\cancel{10 \times}.......... \times 1 \times 1} \times 0.0009765625 \times 1$

$= 1 \times 0.0009765625 \times 1$

$= 0.0009765625$

$\therefore$ Probability of getting atleast heads is

$= 0.044578125 + 0.009765625 + 0.0009765625$

$= 0.05553203125$

(iii) No Heads

$r = 0$

$P(X = 0) = 10_{C_0} \cdot (0.5)^0 \times (0.5)^{10-0}$

$= \frac{10 \times 9 \times ..... \times_1}{1 \times 10 \times ..... \times_1} \times (0.5)^0 \times (0.5)10 = 1 \times 1 \times 0..0009765625$

$P(x = 0) = 0.00097625625$

(iv) At least 1 head

$r \geq 1$ $(r = 1 + r = 2 + r = 3 + ... r = 10)$

(or)

$r \geq 1$ $[1 - (r = 0)]$

$= 1 - 0.0009765625 = 0.999023475.$

(v) Not More Than 3 Heads

$r \leq 3$ $(r = 0 + r = 1 + r = 2 + r = 3)$

$p(r = 1) = 10_{C_1} (0.5)^1 \times (0.5)^{10-1}.$

$= \frac{10 \times \cancel{9 \times .... \times 1}}{1 \times \cancel{9 \times .... \times 1}} \times (0.5)^1 \times (0.5)^9$

$= 10 \times 0.5 \times 0.001953125 = 0.009765625$

$p(r = 2) = 10_{C_2} \times (0.5)^2 \times (0.5)^{10-2}$

$= \frac{10 \times 9 \times \cancel{8 \times ...... \times 1}}{2 \times 1 \times \cancel{8 \times ..... \times 1}} \times (0.5)^2 \times (0.5)8$

$= \frac{90}{2} \times (0.5)^2 \times (0.5)^8$

$= 45 \times 0.25 \times 0.00390625 = 0.0439453125$

$p(r = 3) = 10_{C_3} \times (0.5)^3 \times (0.5)^{10-3}$

$= \frac{10 \times 9 \times 8 \times \cancel{7 \times ..... \times 1}}{3 \times 2 \times 1 \times \cancel{7 \times ..... \times 1}} \times 0.125 \times (0.5)^7$

$= \frac{720}{6} \times 0.125 \times 0.0078125$

$= 120 \times 0.125 \times 0.0078125 = 0.1171875$

$\therefore$ Note more than 3 heads is

$= 0.0009765625 + 0.009765625 + 0.0439453125 + 0.1171875$

$= 0.171872$

(vi) At Least 4 Heads

$r \geq 4$ (r = 4 + r = 5 + r = 6 + ... r = 10)

(or)

1 – [r = 0 + r = 1 + r = 2 + r = 3]

= 1 – [0.0009765625 + 0.009765625 + 0.0439453125 + 0.1171875]

= 1 – 0.171872 = 0.828128.

11. If on an average, rain falls on 12 days in every' 30 days, find the probability (i) that the first 4 days of a given week will be fine, and the remainder wet, (ii) that rain will fall on just 3 days of a given week.

Sol :

The probability of a rainy day, or $p = \frac{12}{30}$, or 2/5

Thus, probability of not rainy day, $q = 1 - \frac{2}{5} = 3/5$

(i) The probability that the first four days will be fine, $q^4 = \left(\frac{3}{5}\right)^4$

The probability that the next 3 days will be wet, $p^3 = \left(\frac{2}{5}\right)^3$

The above two probabilities being independent of each other, their compound probability is given by

$$(q)^4 \times (p)^3 = \left(\frac{3}{5}\right)^4 \times \left(\frac{2}{5}\right)^3$$

$$= \frac{81}{625} \times \frac{8}{125} = \frac{648}{78125}$$

$$= 0.008.$$

(ii) The probability that rain will fall on just 3 days of a given week is given by

$$P(r) = {}^nC_r\, q^{n-r}\, p^r$$

where, n = number of days in a week i.e., 7

r = number of rainy days expected i.e., 3

p = 2/5, and q = 3/5

Thus, $P(3) = {}^7C_3 \left(\frac{3}{5}\right)^4 \left(\frac{2}{5}\right)^3$

$$= 35 \times \frac{81}{625} \times \frac{81}{125} = \frac{22680}{78125} = 0.29$$

Hence, the probability that rain will fall in just 3 days of a given week = 0.29.

12. If on an average, 6 ships out of 10 arrive safely, find the Mean and the S.D. of the number of ships arriving safely out of a total of 1000 ships.

Sol :

For a binomial distribution like the one given, Mean and S.D. are obtained as under.

(i) Mean = np

Where, n = 1C00, and p = probability of safe arrival of a ship i.e., 6/10.

$\therefore$ Mean $= 1000 \times \frac{6}{10} = 600$

(ii) S.D. = $\sqrt{npq}$

Where, $q = 1 - p = 1 - \frac{6}{10} = 4/10$

$$\therefore \sigma = \sqrt{1000 \times \frac{6}{10} \times \frac{5}{10}} = \sqrt{240} = 15.5 \text{ approx.}$$

Hence, the Mean, and the S.D. of the number of ships arriving safely out of 1000 ships are 600 and 15.5 respectively.

13. The normal rate of infection of a certain disease with the children is found to be 20%. On an experiment with 5 children injected with a vaccine, it was observed that none of the children was infected. Calculate the probability of the observed result.

Sol:

Given, the probability of infection, or P = 20% or 1/5

And thus, the probability of non-infection, or q = 4/5.

The size of the sample under study, or n = 5.

The number of children not infected, or r = 0.

Hence, the required probability is given by

$$P_{(r)} = {}^nC_r\, q^{n-r} p^r = {}^5C_0 \left(\frac{4}{5}\right)^5 \left(\frac{1}{5}\right)^0$$

$$= 1 \times \frac{1024}{3125} \times 1 = \frac{1024}{3125} = 0.33 \text{ approx.}$$

$\therefore$ The probability that none of the children will be infected = 0.33, or 1/3.

5.2 Poisson Distribution

5.2.1 Importance

Q6. Define Poisson Distribution. State the importance of Poisson Distribution.

Ans :

Poisson distribution is a discrete probability distribution and is very widely used in statistical work. It was developed by a French mathematician. Simeon Denis Poisson (1781-1840). in 1837. Poisson distribution may be expected in cases where the chance of any individual event being a success is small. The distribution is used to describe the behaviour of rare events such as the number of accidents on road, number of printing mistakes in a book etc., and has been called "the law of improbable events".

The average or mean of Poisson distribution is given by λ. However, the single parameter of Poisson distribution is also given as λ.

Poisson distribution can be used generally to approximate the binomial distribution when 'n' value is large (i.e., n $\rightarrow \infty$) and 'p' value is very small (i.e., p$\rightarrow$0).

Always the sum of infinite probabilities in poisson distribution is 1 i.e.,

$$P(0) + P(1) + P(2) + ... + P(\infty) = 1$$

Importance

1. It is used in quality control statistics to count the number of defects of an item.
2. In biology to count the number of bacteria.
3. In physics to count the number of particles emitted from a radio-active substance.
4. In insurance problems to count the number of casualties.
5. In waiting-time problems to count the number of incoming telephone calls or incoming customers.
6. Number of traffic arrivals such as trucks at terminals, aeroplanes at airports, ships at docks, and so forth.
7. In determining the number of deaths in a district in a given period, say, a year, by a rare disease.
8. The number of typographical errors per page in typed material, number of deaths as a result of road accidents, etc..
9. In problems dealing with the inspection of manufactured products with the probability that any one piece is defective is very small and the lots are very large, and
10. To model the distribution of the number of persons joining a queue (a line) to receive a service or purchase of a product.

Q7. What are the features and assumptions of Poisson distribution?

Ans : **(Imp.)**

Features

(i) It is a discrete probability distribution, since it is concerned with the occurrences that can take only integral values like 0, 1, 2, ... ∞.

(ii) It is applicable to a problem where the probability of happening of an event (p) is very small, that of not happening (q) is not known, and almost equal to unity, n is infinitely large, and the Mean (m) remains constant from trial to trial.

(iii) It has a single parameter m (Mean), with which only all the possible probabilities of the Poisson distribution can be easily obtained. Further, as this parameter m increases, the distribution shifts to the right.

(iv) It is a reasonable approximation of the binomial distribution where, and m is finite.

Assumptions

The theory of Poisson distribution elucidated as above is based on certain assumptions cited as under:

(i) The happening, or non-happening of any event does not affect the happening, or non-happening of any other event.

(ii) The probability of happening of more than one event in a very small interval is negligible.

(iii) The probability of a success for a short time interval or for a short space is proportional to the length of the time interval, or space interval as the case may be.

Q8. What are the applications of Poisson Distribution.

Ans :

i) It is used in the field of physics to find out the number of particles emitted from a 1/ radioactive substance.

ii) It is used in biology to count the number of bacteria in a unit cell.

iii) It is used in the statistical quality control to count the number of defects with a product.

iv) It is used in insurance problems to determine the number of casualties.

v) It is used in the queuing problems to ascertain the number of incoming customers.

vi) It is used in transport and assignment problems to count the number of traffic arrivals such as the trucks at terminals, the Aero planes at airports, the ships at docks etc.

vii) It is used in printing presses to determine the number of typographical errors per page on a typed paper, or the number of printing mistakes per page in a book.

viii) It is used in birth and death registration offices to determine the number of deaths in a particular locality in a given period by a rare disease.

ix) It is used in telephone offices to count the number of telephone-calls arriving at a telephone switch-board per month.

x) It is used to count the number of suicides reported in a particular day, or the number of casualty due to a rare disease such as heart attack, cancer, or snake bite in a year.

xi) It is used in counting the number of accidents taking place per day on a busy road.

xii) It is, also, used in modeling the distribution of number of persons waiting in a line to receive some services.

5.2.2 Conditions and Constants of Poisson Distribution

Q9. Write about Conditions and Constants of Poisson Distribution.

Ans :

Conditions Under Which Poisson Distribution is Used

Poisson distribution is a limiting case of binomial distribution when

1. $n \to \infty$ i.e., number of trials is very large.
2. $P \to 0$ i.e., probability of success for each trial is very small.
3. np = A, is a finite constant (positive real number)

$$\text{Thus, } p = \frac{\lambda}{n} \text{ and } q = \left(1 - \frac{\lambda}{n}\right)$$

Probability of 'x' success in a series of 'n' independent trials is,

$$P(x) = \boxed{P(x) = \frac{e^{-\lambda}.\lambda^{x}}{x!}}$$

Where,

x = 0, 1, 2, ... , n

λ = Mean of poisson distribution.

Constants or General Formula of Poisson Distribution

The probability of 'X' occurrences in poisson distribution is given by,

$$\boxed{P(X) = \frac{e^{-\lambda}.\lambda^{x}}{x!}}$$

Where,

x = Random variable.

It may take 0, 1, 2, 3,...., ∞

e = 2.7183

(The base of natural logarithms)

λ = Mean of poisson distribution

(Average number of occurrences of an event)

Mean can be calculated by multiplying 'n' and 'p'

$$\boxed{\lambda = np}$$

'λ' is the single parameter of Poisson distribution. As 'λ' increases, the distribution shifts to right.

Poisson distribution is a discrete probability distribution with single parameter 'λ'.

As the value of 'λ' increases the distribution shifts to the right. All poisson probability distribution are skewed to right.

Poisson distribution is a distribution of rare events. Here the finite number of trials i.e., the value of 'n' is not mentioned.

'n' tends to oo in this case.

Under Poisson distribution,

$$\boxed{\text{Mean} = \text{Variance} = \lambda}$$

5.2.3 Fitting of Poisson Distribution

Q10. Enumerate the steps involved in Fitting of Poisson Distribution.

Ans :

Following are the steps for fitting a Poisson distribution,

Step 1

Calculate the values of mean (X) and probability of zero occurrence.

Mean in poisson distribution is calculated as,

$\lambda = n \times p$

Where,

n = Number of trials (very large)

P = Probability of successes (very small).

Step 2

Calculate the probabilities by using recurrence relation which is given below.

$$P(x) = \frac{e^{-\lambda}.\lambda^x}{x!} \Rightarrow P(0) = \frac{e^{-\lambda}.\lambda^0}{0!} = e^{-\lambda}$$

Always probability of zero occurrence,

$$P(0) = e^{-\lambda}$$

$$P(1) = \frac{P(0).\lambda}{1} = e^{-\lambda}.\lambda$$

$$P(2) = \frac{P(1).\lambda}{2}$$

$$P(3) = \frac{P(2).\lambda}{3} \text{ and so on.}$$

Step 3

Multiply each term of probabilities with total frequency (TV) to obtain expected frequency values.

PROBLEMS

14. Fit a poisson distribution to the following data.

x:	0	1	2	3	4
f:	123	59	14	3	1

(e^{-m} = 0.6065).

Sol : **(Jan.-21)**

x	f	fx
0	123	0
1	59	59
2	14	28
3	3	9
4	1	4
	200	100

$$\lambda = \frac{\Sigma fx}{\Sigma f} \qquad \lambda = \frac{100}{200} = 0.50$$

Calculation of all probabilities

$$P(0) = 0.6065$$

$$P(1) = \frac{P(0) \times \lambda}{1} = 0.6065 \times 0.5 = 0.3032$$

$$P(2) = \frac{P(1)\times\lambda}{2} = \frac{0.3032\times 0.5}{0.2} = 0.0758$$

$$P(3) = \frac{P(2)\times\lambda}{3} = \frac{0.0758\times 0.5}{3} = 0.013$$

$$P(4) = \frac{P(3)\times\lambda}{4} = \frac{0.013\times 0.5}{4} = 0.002$$

Computation of expected frequencies

x	p(x)	f(x)
0	p(0) = 0.6065	f(0) = 200 × 0.6065 = 121
1	p(1) = 0.3032	f(1) = 200 × 0.3032 = 61
2	p(2) = 0.08	f(2) = 200 × 0.0758 = 15
3	p(3) = 0.013	f(3) = 200 × 0.013 = 3
4	p(4) = 0.001	f(4) = 200 × 0.001 = 0
		200

15. The following mistakes per page were observed in a book:

No. of mistakes per page	0	1	2	3	4
No. of times the mistake occurred	**211**	**90**	**19**	**5**	**0**

Fit a Poisson distribution to fit data.

Sol : **(June-19)**

X	f	fX
0	211	0
1	90	90
2	19	38
3	5	15
4	0	0
	N = 325	ΣfX = 143

$$\bar{X} = \frac{\Sigma f X}{N}$$

$$= \frac{143}{325} = 0.44$$

Mean of the distribution or

m = .44

$(P_0) = e^{-m} = 2.7183^{-.44}$

$= \text{Rec. [antilog (.44 log 2.7183)]}$

$= \text{Rec. [antilog } (.44 \times .4343)]$

$= \text{Rec. [antilog 1.1910]}$

$= \text{Rec. } 1.552 = .6433$

$N(P_0) = .6443 \times 325 = 209.40$

$N(P_1) = N(P_0) \times \frac{m}{1} = 209.4 \times .44 = 92.14$

$N_2(P_2) = N(P_1) \times \frac{m}{2} = 92.14 \times \frac{.44}{2} = 92.14 \times .22 = 20.27$

$N(P_3) = N(P_2) \times \frac{m}{3} = 20.27 \times \frac{.44}{3} = 6.76 \times .44 = 2.97$

$N(P_4) = N(P_3) \times \frac{m}{4} = 2.97 \times \frac{.44}{4} = 2.97 \times .11 = 0.33.$

The expected frequencies of Poisson distribution are :

X	0	1	2	3	4	
f	209.40	92.14	20.27	2.97	0.33	= 325.14

Note :

A rough cehck on the accuracy of result is that the total of the expected frequencies should be equal to the total of the observed frequencies. For example, in the above case the total of expected frequencies is 325.14 and the observed total is 325. The slight difference is due to approximation.

16. The number of defects per unit in a sample of 330 units of manufactured product was found as follows:

No. of defects :	**0**	**1**	**2**	**3**	**4**
No. of units :	**214**	**92**	**20**	**3**	**1**

Fit a Poisson distribution to the data and test for goodness of fit. (Given $e^{-0.439} = 0.6447$).

Sol :

Fitting of Poisson Distribution

X	f	f X
0	214	0
1	92	92
2	20	40
3	3	9
4	1	4
	N = 330	ΣfX = 145

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{145}{330} = 0.439$$

$$P(0) = e^{-m} = e^{-0.439} = 0.439$$

$$N(P_0) = P_{(0)} \times N = .6447 \times 330 = 212.75$$

$$N(P_1) = N(P_0) \times m = 212.75 \times .439 = 93.4$$

$$N(P_2) = N(P_1) \times \frac{m}{2} = 93.4 \times \frac{.439}{3} = 20.5$$

$$N(P_3) = N(P_2) \times \frac{m}{3} = 20.5 \times \frac{.439}{3} = 3.0$$

$$N(P_4) = N(P_3) \times \frac{m}{4} = 3.0 \times \frac{.439}{4} = 0.33$$

Thus the expected frequencies as per Poisson distribution are:

No. of defects	0	1	2	3	4
No. of units	212.75	93.40	20.50	3.00	0.33

17. The following table gives the number of days in a 50 day period during which automobile accidents occurred in a certain part of a city. Fit a Poisson distribution to the data.

No. of accidents	**0**	**1**	**2**	**3**	**4**
No. of days	**19**	**18**	**8**	**4**	**1**

Sol :

Fitting of Poisson Distribution

X	f	fx
0	19	0
1	18	18
2	8	16
3	4	12
4	1	4
	N = 50	ΣfX = 50

$$m = \frac{\Sigma f X}{N} = \frac{50}{50} = 1$$

$$(P_0) = e^{-m} = 2.7183^{-1} = 0.36788 \text{ (from the table)}$$

$$N(P_0) = P_{(0)} \times N = .36788 \times 50 = 18.394 \text{ or } 18.4$$

$$N(P_1) = N(P_0) \times m = 18.394 \times 1 = 18.394 \text{ or } 18.4$$

$$N(P_2) = N(P_1) \times \frac{m}{2} = \frac{18.394}{2} = 9.197 \text{ or } 9.2$$

$$N(P_3) = N(P_2) \times \frac{m}{3} = \frac{9.197}{3} = 3.066 \text{ or } 3.1$$

$$N(P_4) = N(P_3) \times \frac{m}{4} = \frac{3.066}{4} = 0.766 \text{ or } 0.8$$

5.3 Normal Distribution

5.3.1 Importance

Q11. What is Normal Distribution? Explain the uses of normal distribution.

Ans :

Introduction

The Normal Distribution was discovered by De Moivre as the limiting case of Binomial model in 1733. It was also known to Laplace no later than 1774, but through a historical error it has been credited to Gauss who first made reference to it in 1809. Throughout the 18th and 19th centuries, various efforts were made to establish the Normal model as the underlying law ruling all continuous random variables – thus the name Normal. The Normal model has, nevertheless, become the most important probability model is statistical analysis.

The normal Distribution in approximation to Binomial Distribution, whether or not p is equal to q, the Binomial Distribution tends to the form of the continuous cure when n becomes large at least for the material part of the range. As a matter of fact, the correspondence between Binomial and the Normal curve is surprisingly close even for low values of n provide dp and q are fairly near to equality. The limiting frequency curve, obtained as n, becomes large and is called the Normal frequency curve or simply the Normal curve.

Uses of Normal Distribution

i) The normal distribution can be used to approximate Binomial and Poisson distributions.

ii) It has extensive use in sampling theory. It helps us to estimate parameter from stastistic and to find confidence limits of the parameter.

iii) It has a wide use in testing Statistical Hypothesis and Tests of significance in which it is always assumed that the population from which the samples have been drawn should have normal distribution.

iv) It serves as a guiding instrument in the analysis and interpretation of statistical data.

Q12. What are the applications of normal distribution?

Ans :

i) Data obtained from Psychological, Physical and Biological measurements approximately follow Normal distribution. I.Q. scores, heights and weights of individuals etc., are examples of measurements which are normally distributed or nearly so.

ii) Most of the distributions that are encountered in practice, for example, Binomial, Poisson, Hypergeometric, etc. can be approximated to Normal distribution. If the number of trials *n* is indefinitely large and neither *p* nor *q* is very small, then Binomial distribution tends to Normal distribution. If the parameter $\lambda \to \infty$, then Poisson distribution tends to Normal distribution.

iii) Since the Normal distribution is a limiting case of the Binomial distribution for exceptionally large numbers, it is applicable to many applied problems in kinetic theory of gases and fluctuations in the magnitude of an electric current.

iv) Even if a variable is normally distributed, it can sometimes be brought to normal form by simple transformation of the variable.

v) For large samples, any statistic (i.e,, sample mean, sample S.D., etc.) approximately follows Normal distribution and as such it can be studied with the help of normal curve.

vi) Normal curve is used to find confidence limits of the population parameters.

vii) The proofs of all the tests of significance in sampling are based upon the fundamental assumption that the population fronj which the samples have been drawn is normal.

viii) Normal distribution finds large applications in Statistical Quality Control in industry for finding control limits?

Q13. Explain the importance of normal distribution.

Ans :

i) It possesses a lot of mathematical properties for which it is extensively used in a wide variety of fields of physical, natural and social sciences for making various types of analysis,

ii) It is highly useful in statistical quality control in industries for setting up of the control limits.

iii) It is very much used in finding out the probabilities of various sample results as the sample Mean has a theoretical property.

iv) It is highly used in large sampling theory to find out the estimates of the parameters from statistics and confidence limits.

v) It is confirmed by almost all the exact sampling distributions viz : Chi-squire distribution, *t*- distribution, F-distribution, Z-distribution etc for large degree of freedom.

vi) According to the central limit theorem, the normal distribution is used to draw inferences about a universe through sample studies. With this, we can also, estimate the upper and lower limits within which a value in the universe would lie.

vii) This distribution can be regarded as a limiting case of the Poisson distribution, if the value of Mean(m) is infinitely large.

viii) This distribution can be regarded as a limiting case of the binomial distribution, if the value of V (number of trials) is very large, and neither *p* nor *q* is very small.

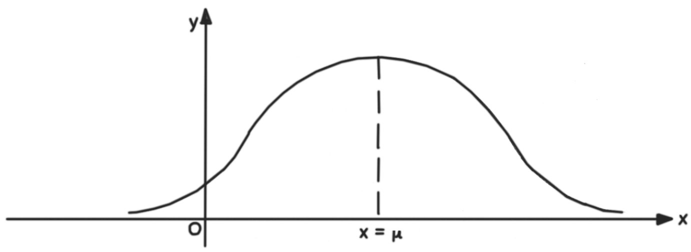
ix) When *n* is very large, computation of probability for most of the discrete distributions e.g. binomial, Poisson etc. becomes quite arduous, and time taking. In such cases normal distribution can be advantageously used with great ease and convenience.

5.3.2 Characteristics

Q14. What are the chief characteristics of normal distribution?

Ans :

i) The graph of the Normal distribution y = f (x) in the xy-plane is known as the N normal curve.



ii) The curve is a bell shaped curve and symmetrical with respect to mean *i.e.,* about the line x = p and the two tails on the right and the left sides of the mean (p) extends to infinity. The top of the bell is directly above the mean p.

iii) Area under the normal curve represents the total population.

iv) Mean, median and mode of the distribution coincide at x = p as the distribution is symmetrical. So normal curve is unimodal (has only one maximum point).

v) x-axis is an asymptote to the curve.

vi) Linear combination of independent normal variates is also a normal variate.

Q15. What are the properties of normal distribution?

Ans : **(June-18, Imp.)**

i) The normal curve is symmetrical about the mean m.

ii) The mean is at the middle and divides the area into halves.

iii) The total area under the curve is equal to 1.

iv) It is completely determined by its mean and standard deviation (s).

5.3.3 Central Limit Theorem

Q16. Explain in detail about Central Limit Theorem.

Ans : **(Imp.)**

The central limit theorem states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough. The shape of the underlying population. The more closely the original population resembles a normal distribution, the fewer sample points will be required.

In practice, some statisticians say that a sample size of 30 is large enough when the population distribution is roughly bell-shaped. Others recommend a sample size of at last 40. But if the original population is distinctly not normal (e.g., is badly skewed, has multiple peaks, and/or has outliers), researchers like the sample size to be even larger.

Following are some of the cases of Central Limit Theorem (C.L.T).

1. De-Moivre's Laplace Theorem

It is the first case of central limit theorem which was stated by Laplace. According to this theorem, the distribution of random variables with respect to the probability of success (P) is asymptotically normal as n tends to infinity.

It can be written as if a random variable,

$$X_i = \begin{cases} 1 & \text{if probability is p} \\ 0 & \text{if probability is q} \end{cases} \quad \text{where } i = 1, 2, 3, ... n$$

Then the distribution Sn = $X_1 + X_2 + X_3 .... X_n$ is normal as $n \to \infty$.

2. Lindeberg-Levy Theorem

This theorem was proposed by Lindeberg and Levy by considering two assumptions.

(i) The distribution of random variables is independent and identical.

(ii) Variance (σ^2) must be finite.

This theorem states that under the above assumptions if the random variables are distributed with $E(X_i)$ = and $V(X_i) = \sigma^2$ then the sum $S_n = X_1+X_2+X_3 + ... + X_n$ follows normal distribution where $\mu^2 = n\sigma_1^2$ (mean) and $\sigma^2 = n\sigma_1^2$ (variance).

3. Liapounoff's Central Limit Theorem

This is a generalized case of central limit theorem where the distribution of random variables is not identical. In this case third absolute moment (p^3) is considered whose distribution can be given as,

$$\rho^3 = \sum_{i=1}^{n} \rho_i^3$$

Then under general conditions, if $E(X_1) = \mu_1$ and $V(X_i) = \sigma_1^2$ and $\underset{n\to\infty}{Lt} = \frac{p}{\sigma} = 0$, then the sum X $= X_1 + X_2 + X_3$ ..... X_n follows normal distribution at $N(\mu_1, \sigma^2)$ with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

5.3.4 Fitting a Normal Distribution (Areas Method Only)

Q17. Write about the area under the normal distribution.

Ans :

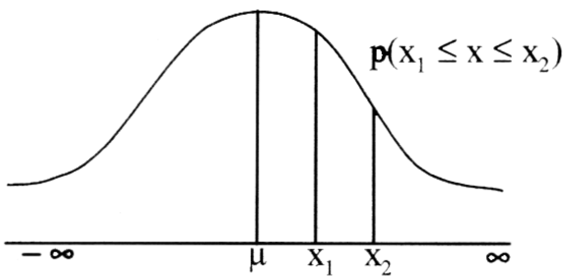
Area under any normal curve is found from the table of standard normal probability distribution showing the area between The mean and any value of the normally distributed random variable.

Since for different values of p and cr we have different normal curves. Hence it is not possible to draw the normal curves for various values of p and a. Thus, the normal curve is transformed into a standardized normal curve.

'x' is transformed into V which is known as 'standard normal variate'.

$$\therefore \quad \boxed{f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}}$$

The area under the normal curve between the ordinates x = 'x_1' and x = 'x_2' gives the probability that the normal variate lies between 'x_1' and 'x_2' as shown in figure.



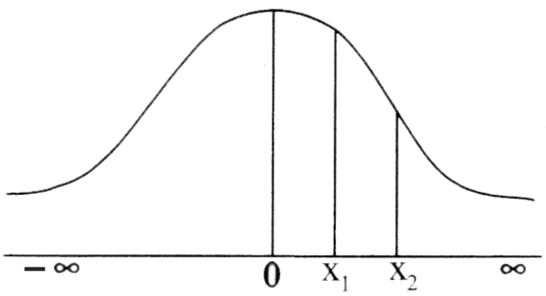
The probability can be evaluated as standardized 'z' as follows,

Substitute, $z = \frac{x-\mu}{\sigma}$ [m = 0, σ = 1]

When, $x = x_1$, $z = \frac{x_1-\mu}{\sigma} = z_1$ (suppose)

When, $x = x_2$, $z = \frac{x_2-\mu}{\sigma}$ = z2

$\therefore \quad p(x_1 \le x \le x_2) = p\,(z_1 \le z \le z_2)$



PROBLEMS

18. For a normally distributed variate with mean 1 and standard deviation 3, find the probabilities that (i) 3.43 $\le$ x $\le$ 6.19 (ii) – 1.43 $\le$ x $\le$ 6.19.

Sol :

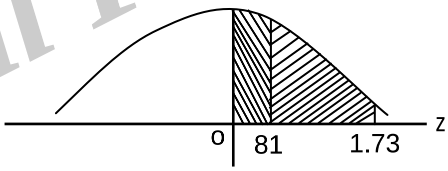
Given $\mu = 1$ and $\sigma = 3$

(i) When x = 3.43,

$$z = \frac{x-\mu}{\sigma} = \frac{3.43-1}{3} = \frac{2.43}{3} = 0.81 = z_1 \text{ (say)}$$

When x = 6.19,

$$z = \frac{x-\mu}{\sigma} = \frac{6.19-1}{3} = \frac{5.19}{3} = 1.73 = z_2 \text{ (say)}$$



$\therefore \quad P(3.43 \le x \le 6.19) = P(0.81 \le z \le 1.73)$

$= |A(z_2) - A(z_1)|$

$= |A(1.73) - A(0.81)|$

$= 0.4582^* - 0.2910$ (from Normal tables)

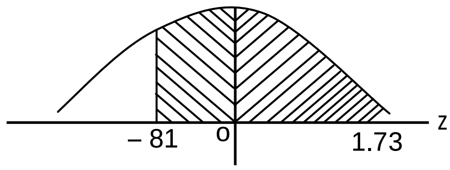
$= 0.1672$ (Cross hatched area in the figure)

(ii) When x = – 1.43,

$$z = \frac{x-\mu}{\sigma} = \frac{-1.43-1}{3} = -0.81 = z_1 \text{ (say)}$$

When x = 6.19,

$$z = \frac{x-\mu}{\sigma} = \frac{6.19-1}{3} = -1.73 = z_2 \text{ (say)}$$



$\therefore$ P(– 1.43 ≤ x ≤ 6.19) = P(– 0.81 ≤ z ≤ 1.73)

= $A(z_2) + A(z_1)$

= A(1.73) + A(–0.81)

= A(1.73) + A(0.81) [$\because$ A(–z) = A(z)]

= 0.4582 + 0.2910 = 0.7492 = shaded area in the figure.

---

19. If X is a normal variate with mean 30 and standard deviation 5. Find (i) P(26 ≤ X ≤ 40) (ii) P(X ≥ 45).

Sol :

Given mean, μ = 30 and S.D., σ = 5

(i) When x = 26, $z = \dfrac{x-\mu}{\sigma} = \dfrac{26-30}{5} = -0.8 = z_1$ (say)

When x = 40, $z = \dfrac{x-\mu}{\sigma} = \dfrac{40-30}{5} = 2 = z_2$ (say)

$\therefore$ P(26 ≤ X ≤ 40) = P(–0.8 ≤ z ≤ 2)

= $A(z_2)$ + A(z1) = A(2) + A(–0.8)

= 0.4772 + 0.2881 = 0.7653 (from normal distribution tables)

(ii) When x = 45, $z = \dfrac{x-\mu}{\sigma} = \dfrac{45-30}{5} = 3$ = z1 (say)

$\therefore$ P(X ≥ 45) = P(z_1 ≥ 3)

= 0.5 – $A(z_1)$ = 0.5 – A(3)

= 0.5 – 0.49865 = 0.00135.

---

20. A study of past participants indicates that the mean length of times spent on the programme is 500 hours; and that, this normal distribution random variable has a standard deviation of 100 hours. What is the probability that a participant selected at random will required to complete the programme in following cases

(i) 'More' than 500 hrs.

(ii) Between 500 and 650 hours.

(iii) Between 550 and 650 hours.

(iv) Less than 580 hours

(v) Between 420 and 570 hours

Sol: **(June -18)**

Let μ be the mean and σ the standard Deviation of the sales. Then we are given that $\mu = 500$ & $\sigma = 100$hrs.

Let the variable 'X' required to complete the programme.

(i) More than 500 hrs

When $X = 500$ Then $Z_1 = \frac{500-500}{100} = 0$ (say Z_1)

$Z_1 = 0$ (From ND table there no value of $Z_1 = 0$)

(ii) Between 500 and 650 hrs

When $X_2 = 500$ then $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{500-500}{100} = 0$ (Say Z_1)

$Z_1 = 0$ (From ND table there no value of $Z_1 = 0$]

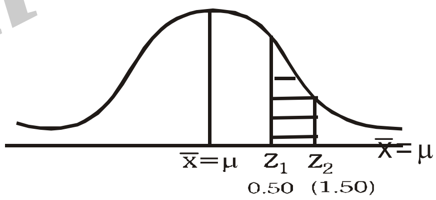
When $X_2 = 650$ then $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{650-500}{100} = \frac{150}{100} = 1.5$ (say Z_2)

$\therefore P(500 \le X \le 650) = P(0 \le Z \le 1.5)$

$= \left|A(Z_2) - A(Z_1)\right|$

$= |0.04332 - 0|$

$= 0.0433.$



(iii) Between 550 and 650 hrs.

When $X_1 = 550$ then $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{550-500}{100} = \frac{50}{100} = 0.50$ (Say Z_1)

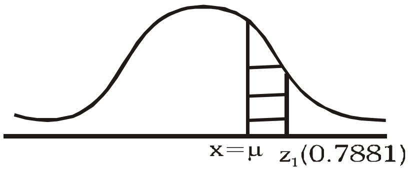
When $X_2 = 650$ then $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{650-500}{100} = \frac{150}{100} = 1.5$ (Say Z_2)

$\therefore P(550 \le x \le 650) = P(0.50 \le z \le 1.5)$

$|A(Z_2) - A(Z_1)|$

$|(0.4332 - 0.1916)|$

$P(550 \le x \le 650) = 0.2416$



(iv) Less than 580 hrs

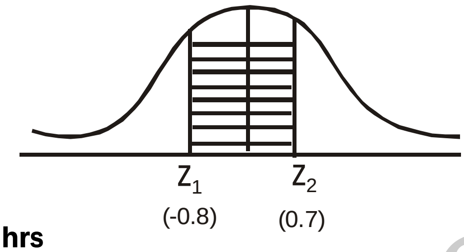
$P(X < 580) = P(Z < Z_1)$

Then X = 580, then $Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{580-500}{100} = \frac{80}{100} = 0.8$(Say Z_1)

$P(X<580) = P(Z<0.8)$

$|0.5 + A(Z_1)| = |0.5 + A(0.8)| = 0.5 + 0.2881$

$P(X< 580) = 0.7881$



(v) Between 420 and 570 hrs

When X_1 = 420 then $Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{420-500}{100} = \frac{-80}{100} = -0.8$(Say Z_3)

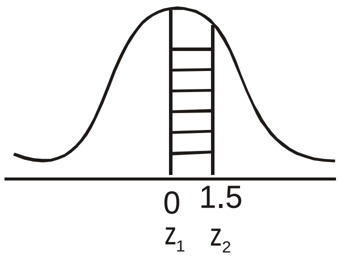
When X_2 = 570 then $Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{570-500}{100} = \frac{70}{100} = 0.7$ (Say Z_2)

$P(X_1 \le X \le X_2) = P(420 \le x \le 570) = P(Z_1 \le Z \le Z_2)$

$|A(Z_2) + A(Z_1)| = |A(0.7) + A(-0.8)|$

$|0.2580+(-0.2881)|$

0.5461



21. The mean height of the students of a certain college is 66" with a standard deviation of 3 inches. How many of the said college consisting of 2000 students would you expect to be over 5 feet height? Give your answer through a normal curve.

Sol :

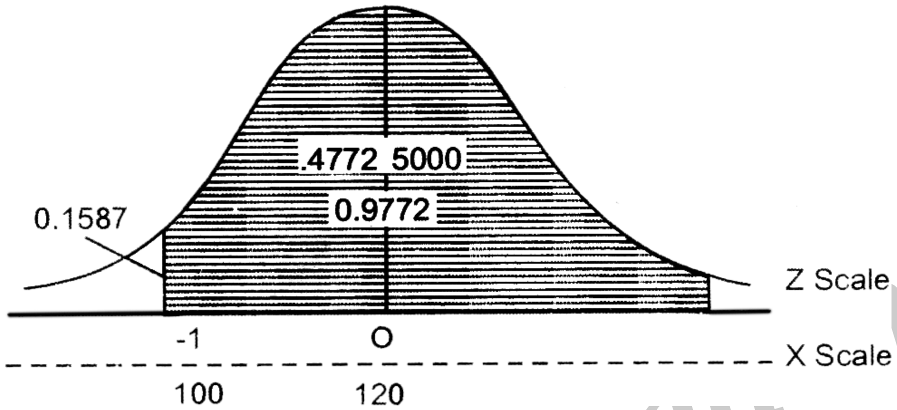
Given, Mean of the distribution, or μ = 66"

Standard deviation, or σ = 3"

Value of the desired height, or X = 5' or 60".

$$\text{Thus, } Z = \frac{X-\mu}{\sigma} = \frac{60-66}{3} = -\frac{6}{3} = -2$$

In a normal curve, the area between Z at 0 and Z at –2 = 0.4772. Further, the area to the of Z at 0 = 0.5000. So, the total area to the of Z at –2 would be 0.4772 + 0.5000 = 0.9772. This is indicated through the shaded portion of a normal curve drawn as under :



Now, the proportion of the area in the normal curve for the students having the height above 5' or 60" = 0.9772

The total number of students or N = 2000

Hence, the number of students expected to have heights above 5 feet would be 0.9772 × 2000 = 1954.4 or 1954.

22. 1000 electric bulbs with a mean life of 120 days are installed in a new factory. Their length of life is normally distributed with a standard deviation of 20 days. How many bulbs will expire in less than 100 days ? If it is decided to replace all the bulbs together, what interval should be allowed between the replacements, if not more than 20% should expire before the replacement ?

Sol :

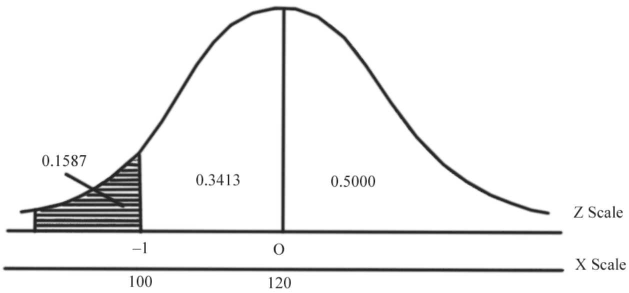
(i) Number of bulbs expected to expire in less than 100 days

Given, μ = 120, σ = 20, value of the desired limit, or X = 100, and N = 1000.

$$\text{Thus, } Z = \frac{X-\mu}{\sigma} = \frac{100-120}{20}$$

$$= -\frac{20}{20} = -1$$

Now, the area of the normal curve less than Z at -1 would be 1 minus the sum of the area between Z at 0 and Z at –1 (i.e. 0.3413 w.r.t. to the Area table), and the area to the of Z at 0 (i.e. 5000). Thus the total area less than Z at –1 = 1 – (0.3413 + 0.5000) = 1– 0.8413 = 0.1587. This is indicated through the shaded portion of a normal curve displayed as under :



Now, the proportion of the area for the bulbs expiring in less than 100 days = 0.1587

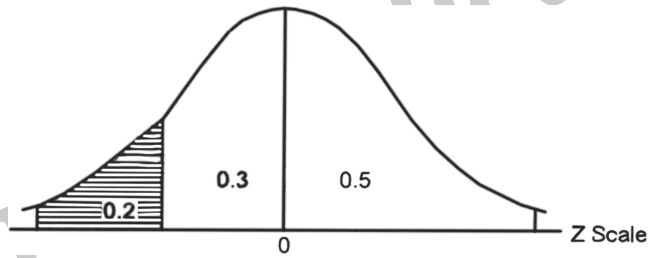
Hence, the number of bulbs expected to expire in less than 100 days = 0.1587 × N

= 0.1587 × 1000 = 158.7 = 159.

(ii) Replacement time allowable

If not more than 20% should expire, then the proportion of the area in the curve for non expiring bulbs would extend upto 0.3 to the of Z at 0 (i.e. 80% – 50% to the of Z at 0).

Thus, the area of 20% representing the expiring bulbs will be the most 0.2 portion of the curve. This is shown through the shaded portion of the normal curve displayed as under.



Now, from the Area Table of the Normal Curve with reference to the area proportion of 0.3 (*i.e.* 0.2995) in a reverse manner we find that Z is at 0.84.

We have, $Z = \frac{X - \mu}{\sigma}$

$\Rightarrow \quad 84 = \frac{X - 120}{20}$

$\Rightarrow \quad 16.8 = X - 120$

$\Rightarrow \quad X = 120 + 16.8 = 136.8 = 137$

Hence, 137 days of interval should be allowed between the replacement, if not more than 20% of the bulbs should expire before replacement.

Short Question and Answers

1. What is Binomial Distribution?

Ans :

Meaning

Binomial distribution was discovered by James Bernoulli in the year 1700 and it is a discrete probability distribution.

The Binomial distribution can be described with the help of a function and is derived as follows:

Let the number of trials be n. The trials be independent i.e., the success or failure at one trial doesnot affect the outcome of the other trials. Thus the probability of success remains the same from trial to trial. Also let 'p' be the probability of success and 'q' be the probability of failure. Then we have $p + q = 1$ or $q = 1 - p$.

One could be interested in finding the probability of getting r successes, which implies getting (n – r) failures.

One of the ways in which one can get r successes and (n – r) failure is to get r continuous successes first and (n – r) successive failures.

The probability of getting such a sequence of r successes and (n-r) failures is, $= p^r q^{n-r}$ (using Multiplication Theorem of Probability)

However, the number of ways in which one can get r successes and n – r failures is ${}^nC_r$ and the probability for each of these ways is $p^r q^{n-r}$.

Thus, if x is the random variable representing the number of successes, the probability of getting r successes and n – r failures, in n trials, is given by the probability function

$$P(x = r) = {}^nC_r\, p^r q^{n-r},\ r = 0, 1, 2, \ldots, n$$

This probability function is popularly known as the Binomial Distribution.

The probabilities of 0, 1, 2, ..., r, ..., n successes are therefore given by

$$q^n,\ {}^nC_1\, p\, q^{n-1},\ {}^nC_2\, p^2 q^{n-2},\ \ldots,\ {}^nC_r\, p^r q^{n-r},\ \ldots,\ p^n.$$

2. What are the conditions of binomial distribution?

Ans :

Binomial Distribution holds under the following conditions :

i) Trials are repeated under identical conditions for a fixed number of times, say *n* times.

ii) There are only two possible outcomes, *e.g.* success or failure for each trial.

iii) The probability of success in each trial remains constant and does not change from trial to trial.

iv) The trials are independent *i.e.,* the probability of an event in any trial is not affected by the results of any other trial.

3. Define Poisson Distribution.

Ans :

Poisson distribution is a discrete probability distribution and is very widely used in statistical work. It was developed by a French mathematician. Simeon Denis Poisson (1781-1840). in 1837. Poisson distribution may be expected in cases where the chance of any individual event being a success is small. The distribution is used to describe the behaviour of rare events such as the number of accidents on road, number of printing mistakes in a book etc., and has been called "the law of improbable events".

The average or mean of poisson distribution is given by λ. However, the single parameter of poisson distribution is also given as λ.

Poisson distribution can be used generally to approximate the binomial distribution when 'n' value is large (i.e., $n \to \infty$) and 'p' value is very small (i.e., $p \to 0$).

Always the sum of infinite probabilites in poisson distribution is 1 i.e.,

$$\boxed{P(0) + P(1) + P(2) + ... + P(\infty) = 1}$$

4. Importance of Poisson Distribution.

Ans :

i) It is used in quality control statistics to count the number of defects of an item.

ii) In biology to count the number of bacteria.

iii) In physics to count the number of particles emitted from a radio-active substance.

iv) In insurance problems to count the number of casualties.

v) In waiting-time problems to count the number of incoming telephone calls or incoming customers.

vi) Number of traffic arrivals such as trucks at terminals, aeroplanes at airports, ships at docks, and so forth.

vii) In determining the number of deaths in a district in a given period, say, a year, by a rare disease

viii) The number of typographical errors per page in typed material, number of deaths as a result of road accidents, etc..

ix) In problems dealing with the inspection of manufactured products with the probability that any one piece is defective is very small and the lots are very large, and

x) To model the distribution of the number of persons joining a queue (a line) to receive a service or purchase of a product.

5. What is Normal Distribution?

Ans :

The Normal Distribution was discovered by De Moivre as the limiting case of Binomial model in 1733. It was also known to Laplace no later than 1774, but through a historical error it has been credited to Gauss who first made reference to it in 1809. Throughout the 18th and 19th centuries, various efforts were made to establish the Normal model as the underlying law ruling all continuous random variables – thus the name Normal. The Normal model has, nevertheless, become the most important probability model is statistical analysis.

The normal Distribution in approximation to Binomial Distribution, whether or not p is equal to q, the Binomial Distribution tends to the form of the continuous cure when n becomes large at least for the material part of the range. As a matter of fact, the correspondence between Binomial and the Normal curve is surprisingly close even for low values of n provide dp and q are fairly near to equality. The limiting frequency curve, obtained as n, becomes large and is called the Normal frequency curve or simply the Normal curve.

6. What are the applications of normal distribution?

Ans :

i) Data obtained from Psychological, Physical and Biological measurements approximately follow Normal distribution. I.Q. scores, heights and weights of individuals etc., are examples of measurements which are normally distributed or nearly so.

ii) Most of the distributions that are encountered in practice, for example, Binomial, Poisson, Hypergeometric, etc. can be approximated to Normal distribution. If the number of trials *n* is indefinitely large and neither *p* nor *q* is very small, then Binomial distribution tends to Normal distribution. If the parameter $\lambda \to \infty$, then Poisson distribution tends to Normal distribution.

iii) Since the Normal distribution is a limiting case of the Binomial distribution for exceptionally large numbers, it is applicable to many applied problems in kinetic theory of gases and fluctuations in the magnitude of an electric current.

iv) Even if a variable is normally distributed, it can sometimes be brought to normal form by simple transformation of the variable.

v) For large samples, any statistic (i.e,, sample mean, sample S.D., etc.) approximately follows Normal distribution and as such it can be studied with the help of normal curve.

Q7. What are the properties of normal distribution?

Ans :

i) The normal curve is symmetrical about the mean m.

ii) The mean is at the middle and divides the area into halves.

iii) The total area under the curve is equal to 1.

iv) It is completely determined by its mean and standard deviation (s).

8. What are the chief characteristics of normal distribution?

Ans :

i) The curve is a bell shaped curve and symmetrical with respect to mean *i.e.,* about the line x = p and the two tails on the right and the left sides of the mean (p) extends to infinity. The top of the bell is directly above the mean p.

i) Area under the normal curve represents the total population.

iii) Mean, median and mode of the distribution coincide at *x* = p as the distribution is symmetrical. So normal curve is unimodal (has only one maximum point).

iv) x-axis is an asymptote to the curve.

v) Linear combination of independent normal variates is also a normal variate.

9. Uses of Normal Distribution

Ans :

i) The normal distribution can be used to approximate Binomial and Poisson distributions.

ii) It has extensive use in sampling theory. It helps us to estimate parameter from stastistic and to find confidence limits of the parameter.

iii) It has a wide use in testing Statistical Hypothesis and Tests of significance in which it is always assumed that the population from which the samples have been drawn should have normal distribution.

iv) It serves as a guiding instrument in the analysis and interpretation of statistical data.

10. Importance of Binomial Distribution.

Ans :

The binomial probability distribution is a discrete probability distribution that is useful in describing an enormous variety of real life events. For example, a quality control inspector wants to know the probability of defective light bulbs in a random sample of 10 bulbs if 10 per cent of the bulbs are defective. He can quickly obtain the answer from tables of the binomial probability distribution. The binomial distribution can be used when :

i) The outcome or results of each trial in the process are characterized as one of two types of possible outcomes. In other words, they are attributes.

ii) The possibility of outcome of any trial does not change and is' independent of the results of previous trials.

Exercise Problems

1. In Delhi with 100 municipal wards, each having approximately the same population, the distribution of Meningitis it is cases in 2007 was as follows:

No. of cases:	0	1	2	3	4
No. of wards:	63	28	6	2	1

Fit a Poisson Distribution for the above.

(Ans: 60.64, 30.32, 7.58, 1.26, 0.16)

2. The distribution of typing mistakes committed by a data entry operator is given below. Assuming a Poisson Model, find out the expected frequencies.

Mistakes per page:	0	1	2	3	4	5
No. of pages:	142	156	69	27	5	1

(Ans: 147.16, 147.16, 73.58, 24.52, 6.13, 1.22)

3. Fit a Poisson Distribution to the following data:

Number of mistakes per page:	0	1	2	3	4	Total
No. of pages:	109	65	22	3	1	200

(Ans: 108.64, 66.27, 20.21, 4.11, 0.63)

4. For a Binomial distribution, the mean and variance are respectively 4 and 3. Calculate the probability of getting a non-zero value of this distribution.

(Ans : $1-\left(\frac{3}{4}\right)^{16}$)

5. In a Binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 an 0.2048 respectively. Find the parameter 'p' of the distribution.

(Ans: p = 0.2)

6. A's chance of winning a single game against B is $\frac{2}{3}$. Find the chance that in a series of 5 games with B, A wins: (i) exactly three games, (ii) at least three games:

(Ans: (i) $\frac{80}{243}$, (ii) $\frac{64}{81}$)

7. The customer accounts at a certain departmental store have an average balance of ` 480 and a standard deviation of ` 160. Assuming that the account balance are normally distributed.

(i) What proportion of the accounts is over ` 600?

(ii) What proportion of the accounts is between ` 400 and ` 600?

(iii) What proportion of the accounts is between ` 240 and ` 360?

(Ans: (i) 22.66% (ii) 0.4649% (iii) 15.98%)

8. The marks obtained in a certain examination follow the normal distribution with mean 45 and standard Deviation 10. If 1000 students appeared in the examination, calculate the number of students scoring:

   (i) Less than 40 marks,

   (ii) More than 60 marks ,

   (iii) Between 40 and 50 marks.

   (Ans: (i) 309 (ii) 67 (iii) 383)

9. In a manufacturing organization, the distribution of wages was perfectly normal and the number of workers employed in the organization was 5000. The mean wages of the workers were calculated as ` 800 per day and the standard deviation was worked out to be ` 200. On the basis of the information, estimate:

   (i) The number of workers getting wages between ` 700 and ` 900.

   (ii) Percentage of workers getting wages above ` 1000.

   (iii) Percentage of workers getting wages below ` 600.

   (Ans: (i) 1915 (ii) 15.87% (iii) 15.87%)

10. The average selling price of houses in a city is ` 50,000 with a standard deviation of ` 10,000. Assuming the distribution of selling price to be normal find: (i) The percentage of houses that sell for more than ` 55000 (ii) the percentage of houses selling between ` 45000 and ` 60,000.

    (Area between Z = 0 and Z = 1 is = 0.3413 and the area between Z = 0 and Z = is 0.1915, where Z is a standard normal variate)

    (Ans: (i) 30.85% (ii) 53.28%)

Choose the Correct Answer

1. Which of the following is not a condition of the binomial distribution [c]
   (a) Only 2 possible outcomes
   (b) have constant probability of success
   (c) must have at least 3 trials
   (d) None
2. A variable that can assume any possible value between two points is called: [b]
   (a) Discrete random variable
   (b) Continuous random variable
   (c) Discrete sample space
   (d) Random variable
3. A discrete probability distribution may be represented by: [d]
   (a) Table (b) Graph
   (c) Mathematical equation (d) All of the above
4. If C is a constant in a continuous probability distribution, then p(x = C) is always equal to: [a]
   (a) Zero (b) One
   (c) Negative (d) Impossible
5. If the random variable takes negative values, then the negative values will have: [a]
   (a) Positive probabilities (b) Negative probabilities
   (c) Constant probabilities (d) None
6. Which one is not an example of random experiment? [d]
   (a) A coin is tossed and the outcome is either a head or a tail
   (b) A six-sided die is rolled
   (c) Some number of persons will be admitted to a hospital emergency room during any hour.
   (d) All medical insurance claims received by a company in a given year.
7. The probability function is always [c]
   (a) Negative (b) Positive
   (c) Non negative (d) None
8. Area under the normal curve on either side of mean is [a]
   (a) 0.5 (b) 1
   (c) 2 (d) None
9. Normal Distribution is [a]
   (a) Mesokurtic (b) Leptokurtic
   (c) Platykurtic (d) None
10. Which of the following parameter control the relative flatness of normal distribution [a]
   (a) Standard Deviation (b) Mean
   (c) Mode (d) None

Fill in the blanks

1. A single-equation econometric model of the demand for a product is a ________ equation in which the quantity demanded of the product is an ________ variable.
2. Trend projection is an example of ________.
3. Turning points in the level of economic activity can be forecast by using ________.
4. The graph of time series is called________.
5. Secular trend can be classified in to ________.
6. In time series seasonal variations can occur within a period of ________.
7. Time-series analysis is based on the assumption that ________.
8. A single-equation econometric model of the demand for a product is a ________ equation in which the quantity demanded of the product is an ________ variable
9. Barometric methods are used to forecast ________.
10. A ________ is a statistical data that are collected, observed or recorded at regular intervals of time.
11. The probability of getting 2 heads in tossing 5 coins is ________.
12. If a coin is tossed 6 times in succession, the probability of getting at least one head is ________.
13. The mean of the binomial distribution is ________.
14. If n and p are the parameters of a binomial distribution, the standard deviation of this distribution is________.
15. The probability of having at least one tail in four throws with a coin is ________.
16. If mean of the binomial distribution is 8 and variance is 6, the mode of this distribution is________.
17. If mean of the binomial distribution is 6 and variance is 2, then mode of this distribution is ________.
18. If mean of the binomial distribution is 4 and variance is 2 then p =________.
19. The probability of getting one boy in a family of 4 children is ________.
20. If the mean and variance of a binomial variate are 12 and 4, then the distribution is 25 ________.
21. In a binomial distribution the sum and product of the mean and variance are $\frac{25}{3}$ and $\frac{50}{3}$ respectively. The distribution is________.
22. If the porobability of a defective bolt is 0.1, the mean and standard deviation for the distribution bolts in a total of 400 are ________ and ________ respectively.
23. A coin is tossed 3 times. The probability of obtaining two heads will be ________.

24. The probability of producing a defective bolt is 0.1. The probability that out of 5 bolts one will be defective is __________.

25. The binomial distribution whose mean is 5 and variance is $\frac{10}{3}$ is __________.

26. The probability of getting four heads in six tosses of a fair coin is__________.

27. A die is thrown 8 times. The probability that 3 will show exactly 2 times is__________.

28. The mean, median and mode of a Normal distribution are__________.

29. In the standard normal curve the area between $z = -1$ and $z = 1$ is nearly __________.

30. If $\mu = 5$ and $\sigma = 2$, the equation of the nromal distribution is __________.

Answers

1. Structural, endogenous
2. Time-series
3. Barometric methods
4. Histogram
5. Four methods
6. One year
7. Past patterns in the variable to be forecast will continue unchanged into the future
8. Structural, endogenous
9. Cyclical variation
10. Time series
11. $\frac{5}{256}$
12. $\frac{63}{64}$
13. np
14. $\sqrt{npq}$
15. $\frac{15}{16}$
16. 8
17. 6
18. $\frac{1}{2}$

19. $\frac{1}{4}$

20. $\left(\frac{1}{3}+\frac{2}{3}\right)^{18}$

21. $\left(\frac{2}{3}+\frac{1}{3}\right)^{15}$

22. 40, 6

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

24. $\frac{1}{2}(0.9)^4$

25. $15_{c_r}\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{15-r}$

26. $\frac{15}{64}$

27. $\frac{28\times 5^6}{6^8}$

28. Normal distribution

29. 1

30. 68%

Value of e^x and e^{-x}

x	e^x	e^{-x}	X	e^x	e^{-x}
0.00	1.0000	1.00000	0.40	1.4918	.67032
0.01	1.0101	0.99005	0.41	1.5068	.66365
0.02	1.0202	.98020	0.42	1.5220	.65705
0.03	1.0305	.97045	0.43	1.5373	.65051
0.04	1.0408	.96079	0.44	1.5527	.64404
0.05	1.0513	.95123	0.45	1.5683	.63763
0.06	1.0618	.94176	0.46	1.5841	.63128
0.07	1.0725	.93239	0.47	1.6000	.62500
0.08	1.0833	.92312	0.48	1.6161	.61878
0.09	1.0942	.91393	0.49	1.6323	.61263
0.10	1.1052	.90484	0.50	1.6487	.67053
0.11	1.1163	.89583	0.51	1.6653	.60650
0.12	1.1275	.88692	0.52	1.6820	.59452
0.13	1.1388	.87809	0.53	1.6989	.58860
0.14	1.1503	.86936	0.54	1.7160	.58275
0.15	1.1618	.86071	0.55	1.7333	.57695
0.16	1.1735	.85214	0.56	1.7507	.57121
0.17	1.1853	.84366	0.57	1.7683	.56553
0.18	1.1972	.83527	0.58	1.7860	.55990
0.19	1.2092	.82696	0.59	1.8040	.55433
0.20	1.2214	.81873	0.60	1.8221	.54881
0.21	1.2337	.81058	0.61	1.8404	.54335
0.22	1.2461	.80252	0.62	1.8589	.53794
0.23	1.2586	.79453	0.63	1.8776	.53259
0.24	1.2712	.78663	0.64	1.8965	.52729
0.25	1.2840	.77880	0.65	1.9155	.52205
0.26	1.2969	.77105	0.66	1.9348	.51685
0.27	1.3100	.76338	0.67	1.9542	.511571
0.28	1.3231	.75578	0.68	1.9739	.50662
0.29	1.3364	.74826	0.69	1.9937	.50158
0.30	1.3499	.74082	0.70	2.0138	.49659
0.31	1.3634	.73345	0.71	2.0340	.49164
0.32	1.3771	.72615	0.72	2.0544	.48675
0.33	1.3910	.71892	0.73	2.0751	.48191
0.34	1.4049	.71177	0.74	2.0959	.47711
0.35	1.4191	.70469	0.75	2.1170	.47237
0.36	1.4333	.69768	0.76	2.1383	.46767
0.37	1.4477	.69073	0.77	2.1598	.46301
0.38	1.4623	.68386	0.78	2.1815	.45841
0.39	1.4770	.67706	0.79	2.2034	.45384

x	e^x	e^{-x}
0.80	2.2255	.44933
0.81	2.2479	.44486
0.82	2.2705	.44043
0.83	2.2933	.43605
0.84	2.3164	.43171
0.85	2.3396	.42741
0.86	2.3632	.42316
0.87	2.3869	.41895
0.88	2.4109	.41478
0.89	2.4351	.41066
0.90	2.4596	.40657
0.91	2.4843	.40252
0.92	2.5093	.39852
0.93	2.5345	.39455
0.94	2.5600	.39063
0.95	2.5857	.38674
0.96	2.6117	.38289
0.97	2.6379	.37908
0.98	2.6645	.37531
0.99	2.6912	.37158
1.00	2.7183	.36788
1.10	3.0042	.33287
1.20	3.3201	.30119
1.30	3.6693	.27253
1.40	4.0552	.24660
1.50	4.4817	.22313
1.60	4.9530	.20190
1.70	5.4739	.18268
1.80	6.0496	.16530
1.90	6.6859	.14957
2.00	7.3891	.13534
2.10	8.1662	.12246
2.20	9.2050	.11080
2.30	9.9724	.10026
2.40	11.023	.09072
2.50	12.182	.08208
2.60	13.464	.07427
2.70	14.880	.06721
2.80	16.445	.06081
2.90	18.174	.05502

x	e^x	e^{-x}
3.00	20.086	.04979
3.10	22.198	.04505
3.20	24.533	.04076
3.30	27.113	.03688
3.40	29.964	.03337
3.50	33.115	.03020
3.60	36.598	.02732
3.70	40.447	.02472
3.80	44.701	.02237
3.90	49.402	.02024
4.00	54.598	.01832
4.10	60.340	.01657
4.20	66.686	.01500
4.30	73.700	.01357
4.40	81.451	.01227
4.50	90.107	.01111
4.60	99.484	.01005
4.70	109.95	.00910
4.80	121.51	.00823
4.90	134.29	.00745
5.00	148.41	.00674
5.10	164.02	.00610
5.20	181.27	.00552
5.30	200.34	.00499
5.40	221.41	.00452
5.50	244.69	.00409
5.60	270.43	.00370
5.70	298.87	.00335
5.80	330.30	.00303
5.90	365.04	.00274
6.00	403.43	.00248
6.25	518.01	.00193
6.50	665.14	.00150
6.75	854.05	.00117
7.00	1096.6	.00091
7.50	1808.0	.00055
8.00	2981.0	.00034
8.50	4914.8	.00020
9.00	8103.1	.00012
9.50	13360	.00007
10.00	22026	.00005

BINOMIAL COEFFICIENTS

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	762	924	762	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	3005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27133	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

VALUES OF e–m (For Computing Poisson Probabilities) (0)
(0 < m < 1)

m	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	.9900	.9802	.9704	.9608	.9512	.9418	.9324	.9231	.9139
0.1	0.9048	.8958	.8860	.8781	.8694	.8607	.8521	.8437	.8353	.8270
0.2	0.8187	.8106	.8025	.7945	.7866	.7788	.7711	.7634	.7558	.7483
0.3	0.7408	.7334	.7261	.7189	.7118	.7047	.6977	.6907	.6839	.6771
0.4	0.6703	.6636	.6570	.6505	.6440	.6376	.6313	.6250	.6188	.6126
0.5	0.6065	.6005	.5945	.5886	.5827	.5770	.5712	.5655	.5599	.5543
0.6	0.5488	.5434	.5379	.5326	.5278	.5220	.5160	.5117	.5066	.5016
0.7	0.4966	.4916	.4868	.4810	.4771	.4724	.4670	.4630	.4584	.4538
0.8	0.4493	.4449	.4404	.4360	.4317	.4274	.4232	.4190	.4148	.4107
0.9	0.4066	.4025	.3985	.3986	.3906	.3867	.3839	.3791	.3753	.3716

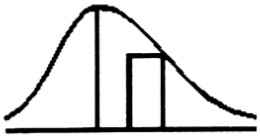
(m : 1, 2, 3 ...10)

m	1	2	3	4	5	6	7	8	9	10
e^{-m}	.36788	.13534	.04979	.01832	.00698	.00279	.00092	.000395	.000123	.000045

Note : To obtain values of e^{-m} for other values of m, use of laws of exponents.

Example. $E^{2.35} = (e^{-2.00} (e^{-0.35} = (.13534)(.7047) = .095374$

ORDINATES (Y) OF THE STANDARD NORMAL CURVE At z



z	0	1	2	3	4	5	6	7	8	9
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	0.3970	0.3965	0.3961	0.3956	0.3951	0.3945	0.3939	0.3932	0.3925	0.3918
0.2	0.3910	0.3902	0.3894	0.3885	0.3876	0.3867	0.3857	0.3847	0.3836	0.3825
0.3	0.3814	0.3802	0.3790	0.3778	0.3765	0.3752	0.3739	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	0.3637	0.3621	0.3605	0.3589	0.3572	0.3558	0.3538
0.5	0.3531	0.3503	0.3485	0.3467	0.3438	0.3429	0.3410	0.3391	0.3372	0.3352
0.6	0.3332	0.3312	0.3292	0.3271	0.3251	0.3230	0.3209	0.3187	0.3166	0.3144
0.7	0.3123	0.3101	0.3079	0.3056	0.3034	0.3011	0.2989	0.2966	0.2943	0.2920
0.8	0.2897	0.2874	0.2850	0.2827	0.2803	0.2780	0.2756	0.2732	0.2709	0.2685
0.9	0.2661	0.2617	0.2613	0.2589	0.2565	0.2541	0.2516	0.2492	0.2468	0.2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	0.2179	0.2151	0.2131	0.2107	0.2083	0.2059	0.2036	0.2012	0.1989	0.1965
1.2	0.1942	0.1919	0.1895	0.1872	0.1849	0.1826	0.1804	0.1781	0.1758	0.1736
1.3	0.1714	0.1691	0.1669	0.1647	0.1626	0.1604	0.1582	0.1561	0.1539	0.1518
1.4	0.1497	0.1476	0.1456	0.1435	0.1415	0.1394	0.1374	0.1354	0.1334	0.1315
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127
1.6	0.1109	0.1092	0.1074	0.1057	0.1040	0.1023	0.1006	0.0989	0.0973	0.0957
1.7	0.0940	0.0925	0.0909	0.0891	0.0878	0.0863	0.0848	0.0833	0.0818	0.0804
1.8	0.0790	0.0775	0.0761	0.0748	0.0734	0.0721	0.0707	0.0694	0.0681	0.0669
1.9	0.0656	0.0644	0.0632	0.0620	0.0608	0.0596	0.0584	0.0573	0.0562	0.0551
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449
2.1	0.0440	0.0431	0.0422	0.0413	0.0404	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	0.0347	0.0339	0.0332	0.0325	0.0317	0.0310	0.0303	0.0297	0.0290
2.3	0.0283	0.0277	0.0270	0.0264	0.0258	0.0252	0.0246	0.0241	0.0235	0.0229
2.4	0.0224	0.0219	0.0213	0.0208	0.0203	0.0198	0.0194	0.0189	0.0184	0.0180
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139
2.6	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.7	0.0104	0.0101	0.0099	0.0096	0.0093	0.0091	0.0088	0.0086	0.0084	0.0081
2.8	0.0079	0.0077	0.0075	0.0073	0.0071	0.0069	0.0067	0.0065	0.0063	0.0061
2.9	0.0060	0.0058	0.0056	0.0055	0.0053	0.0051	0.0050	0.0048	0.0047	0.0046
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034
3.1	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	0.0025	0.0025
3.2	0.0024	0.0023	0.0022	0.0022	0.0021	0.0020	0.0020	0.0019	0.0018	0.0018
3.3	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013	0.0010
3.4	0.0012	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010	0.0010	0.0009	0.0009
3.5	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.0006
3.6	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0004
3.7	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003	0.0003
3.8	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
3.9	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001

FACULTIES OF COMMERCE

B.Com. IV - Semester (CBCS) Examination

(Common Paper for General / Computers and Computer Applications / Advertising / Foreign Trade and tax Procedures)

January - 2021

BUSINESS STATISTICS - II

Time : 2 Hours] **[Max. Marks : 80**

PART - A - (4 × 5 = 20 Marks)

Note : Answer any **FOUR** questions.

ANSWERS

1. If x = 0.85 y and y = 0.89x. Find the coefficient of correlation. **(Unit-I, Prob.2)**

2. Define Index Numbers. **(Unit-II, SQA.3)**

3. From the following data construct on Index Number for 2017 taking 2016 as base as per simple aggregative Method.

Commodity	P	Q	R	S	T
Price 2015 (`)	40	60	85	25	30
Price 2017 (`)	60	90	125	30	40

(Unit-II, Prob.4)

4. Components of Time Series. **(Unit-III, Q.No.4)**

5. From the following data fit a trend line by the method of Semi-Average.

Year :	2012	2013	2014	2015	2016	2017
Output :	20	16	24	30	28	32

(Unit-III, Prob.10)

6. Explain

   (i) Dependent event and **(Unit-IV, SQA.1)**

   (ii) Independent Event. **(Unit-IV, SQA.2)**

7. Explain the Axiomatic Approach to probability. **(Unit-IV, SQA.3)**

8. Comment on the following,

   For a Binomial Distribution mean = 7 and Variance = 11. **(Unit-V, Prob.5)**

PART - B - (4 × 15 = 60 Marks)

Note : Answer any **FOUR** questions.

9. What is meant by regression? What is the importance and limitations of Analysis?

(Unit-I, Q.No.1,2,5)

10. From the following data obtain the two regression equations and calculate the correlation co-efficient.

x	2	4	6	8	10	12	14	16	14
y	18	16	20	24	22	26	28	32	30

Calculate the value of y when x = 6.2 **(Unit-I, Prob.7)**

11. The index of 2010 is 100. It rises by 5% in 2011, falls 2% in 2012, rises 4% in 2013, rises 10% in 2014, falls 3% in 2015, and rises 8% in 2016. Find out the indices for the 7 years assuming that all the increases and decreases are the percentages of the respective proceeding years - Also recast the indices shifting base to 2014. **(Unit-II, Prob.17)**

12. From the following data calculate price index according to

(i) Laspeyre,

(ii) Paasche and

(iii) Marshall-Edgeworth methods.

Item	Base year		Current year	
	Price (`)	Expenditure (`)	Price (`)	Expenditure (`)
A	5	50	8	40
B	7	25	12	30
C	9	10	15	25
D	12	5	20	18

(Unit-II, Prob.11)

13. Calculate the 3 yearly the 5 yearly moving averages for the following time series.

Year :	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Production (in quintals)	500	540	550	530	520	560	600	640	620	610	640

(Unit-III, Prob.16)

14. Obtain the straight line trend equation for the following data by the method of the least square. Tabulate the trend values.

Year :	2010	2011	2012	2013	2014	2015	2016
Sale (in '000 units)	140	144	160	152	168	176	180

(Unit-III, Prob.17)

15. A box contains 8 Red and 5 White balls. Two successive draws of 3 balls are made at random. Find the probability that the first three are white and second three are red.

(i) When there is replacement and

(ii) When there is no replacement. **(Unit-IV, Prob.26)**

16. A factory has two machines. Empirical evidence has established that machines I and II produce 30% and 70% of output respectively. It has also been established that 5% and 1% of the output produced by these machines respectively was defective. A defective item is drawn at random. What is the probability that the defective item was produced by either machine I or machine II ? **(Unit-IV, Prob.30)**

17. 8 Coins are tossed at a time, 256 times. Find the expected frequencies of successes Getting a Head) and tabulate the results obtained. **(Unit-V, Prob.8)**

18. Fit a poisson distribution to the following data.

x :	0	1	2	3	4
f:	123	59	14	3	1

(e^{-m} = 0.6065). **(Unit-V, Prob.14)**

FACULTIES OF COMMERCE

B.Com. IV – Semester (CBCS) Examination

(Common Paper for General / Computers / Computer Applications / Advertising / Foreign Trade / and Tax Procedure Courses)

May/June - 2019

BUSINESS STATISTICS - II

Time : 3 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

Note: Answer any five of the following questions not exceeding 20 lines each.

Answers

1. If r = 0.8; σ_x = 2.5, σ_y = 3.5, find bxy and byx. **(Unit-I, Prob.3)**

2. Types of Index Numbers. **(Unit-II, SQA.2)**

3. Calculate Index number by average price relative method by using arithmetic mean.

Commodity	P	Q	R	S	T
Price 2017	2	6	10	5	12
Price 2018	4	8	15	5	8

(Unit-II, Prob.5)

4. Utility of Time Series Analysis. **(Unit-III, SQA.12)**

5. Fit a trend line to the following data by the freehand method.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Sales (Rs.)	19	22	24	20	23	25	23	26	25

(Unit-III, Prob.9)

6. Explain :

   (i) Mutually Exclusive Events and **(Unit-IV, SQA.4)**

   (ii) Dependent events. **(Unit-IV, SQA.1)**

7. One card is drawn at random from a pack of 52 cards. What is the probability that it is either a king or a queen ? **(Unit-IV, Prob.18)**

8. Comment on the following:

   For a Binomial Distribution Mean = 7 and Variance = 11. **(Unit-V, Prob.5)**

PART - B (5 × 12 = 60 Marks)

Note: Answer all the questions

9. (a) From the following data obtain the two regression equations and calculate the correlation coefficient.

X	2	4	6	8	10	12	14	16	18
Y	18	16	20	24	22	26	28	32	30

(Unit-I, Prob.7)

OR

(b) Following are the marks in Statistics and English in an Annual Examination.

	Statistics (X)	English (Y)
Mean	40	50
Standard Derivation	10	16
Co-efficient Correlation	0.5	

(i) Estimate the score of English, when the score in Statistics is 50.

(ii) Estimate the score of statistics, when the score in English is 30. **(Unit-I, Prob.9)**

10. (a) Define index number. What are its features and uses ? **(Unit-II, Q.No.1,2,3)**

OR

(b) Compute Price Index number by using:

i) Paasches and

ii) Marshal and Edge worth methods.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
P	5	100	6	150
Q	4	80	5	100
R	2	60	5	72
S	12	30	9	33

(Unit-II, Prob.12)

11. (a) From the following data, calculate trend values using Four Yearly Moving Averages.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	506	620	1036	673	588	696	1116	738	663

(Unit-III, Prob.15)

OR

(b) Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

(Unit-IV, Prob.18)

12. (a) A bag contains 4 defective and 6 good Electronic Calculators. Two calculators are drawn at random one after the other without replacement. Find the probability that

i) Two are good

ii) Two are defective and

iii) One is good and one is defective. **(Unit-IV, Prob.19)**

OR

(b) A company has two plants for manufacturing scooters. Plant I manufacturers 80% of the Scooters and Plant II manufactures 20%. At the Plant I 85% Scooters are rated to be of standard quality and at plant II 65% Scooters are rated to be of standard quality. One Scooter was selected at random. What is the probability that

i) It is manufactured by Plant I

ii) It is manufactured by Plant II – which is of standard quality. **(Unit-IV, Prob.31)**

13. (a) Ten unbiased coins are tossed simultaneously. Find the probability of obtaining:

i) Exactly 6 Heads

ii) Atleast 8 Heads

iii) No Heads

iv) Atleast one Head

v) Not more than 3 Heads and

vi) Atleast 4 heads. **(Unit-V, Prob.10)**

OR

(b) Fit a Poisson distribution to the following data:

X	0	1	2	3	4
Y	211	9	19	5	0

(e–m = 0.6443)

(Unit-V, Prob.15)

FACULTIES OF COMMERCE

B.Com. IV – Semester (CBCS) Examination

(Common Paper for General / Computers / Computer Applications / Advertising / Foreign Trade / and Tax Procedure Courses)

May/June - 2018

BUSINESS STATISTICS - II

Time : 3 Hours] **[Max. Marks : 80**

PART - A - (5 × 4 = 20 Marks)

Note : Answer any FIVE of the following questions not exceeding 20 lines each.

ANSWERS

1. If $\gamma = 0.6$, $\sigma x = 1.5$ and $\sigma y = 2$, Find the bxy and byx. **(Unit-I, Prob.4)**
2. Importance of Index Numbers. **(Unit-II, SQA.1)**
3. From the following data calculate a price Index based on price Relatives Method using Arithmetic Mean.

Commodity	A	B	C	D	E	F
Price 2015(Rs.)	45	60	20	50	85	120
Price 2016 (Rs.)	55	70	30	75	90	130

(Unit-II, Prob.3)

4. What are the uses of Time series ? **(Unit-III, SQA.12)**
5. Explain
   (i) Mutually Exclusive Events and
   (ii) Not-Mutually Exclusive Events. **(Unit-IV, SQA.4,5)**
6. When two dice are thrown, find the probability that the sum of the numbers is either 10 or 11. **(Unit-IV, Prob.4)**
7. 6 coins are tossed at a time what is the probability of obtaining 4 or more Heads. **(Unit-V, Prob.4)**
8. Properties of Normal Distribution. **(Unit-V, SQA.7)**

PART - B - (5 × 12 = 60 Marks)

9. (a) Define Regression and what are the differences between correlation and Regression. **(Unit-I, Q.No.1,8)**

OR

   (b) Given :
   $\Sigma x = 56$, $\Sigma y = 40$, $\Sigma x^2 = 524$, $\Sigma y^2 = 256$, $\Sigma xy = 364$, $N = 8$
   (i) Find the two Regression equations and
   (ii) The Correlation Coefficient. **(Unit-I, Prob.8)**

10. (a) The following are the indices (2007. Base)

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Indices	100	120	122	116	120	120	137	136	149	156	137

Shift the base to 2012 and recast the index numbers. **(Unit-II, Prob.18)**

OR

(b) From the following data calculate price index Number by using (i) Paasche's Method and (ii) Method and (ii) Marshal Edgeworth Method.

Item	Base year		Current year	
	Price (Rs.)	Expenditure (Rs.)	Price (Rs.)	Expenditure (Rs.)
P	6	300	10	560
Q	2	200	2	240
R	4	240	6	360
S	10	300	12	288
T	3	120	8	240

(Unit-II, Prob.13)

11. (a) Find the 4 yearly moving averages from the following data.

Year	2008	209	2010	2011	2012	2013	2014	2015	2016	2017
Production (in Tonnes)	150	170	196	180	190	216	248	280	300	320

(Unit-III, Prob.14)

(b) Production figure of a Textile Industry are as follows.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in '000 units)	12	10	14	11	13	15	16

For the above data

(i) Determine the straight line equation under rthe Least Square Method.

(ii) Find the Trend Values and show the trend line on a graph paper. **(Unit-III, Prob.19)**

12. (a) From 30 tickets marked with first 30 numerals, 1 ticket is drawn at random. It is then replaced and a second draw is made. Find the probability that in the first draw it is multiple of 5 or 7 and in second draw it is a multiple of 3 or 7. **(Unit-IV, Prob.27)**

OR

(b) In a bolt factory, the Machines P,Q and R manufacture respectively 25%, 35% and 40% of the total of their outputs 5,4,2 percents respectively are defective bolts. A bolt is drawn at random from the product, and is known to be defective, What are the probabilities that it was manufactured by the machines P,Q and R. **(Unit-IV, Prob.31)**

13. (a) Five coins are tossed 3,200 times, find the frequencies of the distribution of heads and tails, and tabulate the results and also calculate Mean and standard Deviation of fitted distribution. **(Unit-V, Prob.9)**

OR

(b) A study of past participants indicates that the mean length of times spent on the programme is 500 hours; and that, this normal distribution random variable has a standard deviation of 100 hours. What is the probability that a participant selected at random will required to complete the programme in following cases

(i) 'More' than 500 hrs.

(ii) Between 500 and 650 hours.

(iii) Between 550 and 650 hours.

(iv) Less than 580 hours

(v) Between 420 and 570 hours **(Unit-V, Prob.20)**

FACULTIES OF COMMERCE

B.Com. IV – Semester (CBCS) Examination

(Common Paper for General / Computers / Computer Applications / Advertising / Foreign Trade / and Tax Procedure Courses)

MODEL PAPER - I

BUSINESS STATISTICS - II

Time : 3 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

Note: Answer any five of the following questions not exceeding 20 lines each.

ANSWERS

1. Importance of regression analysis **(Unit-I, SQA.2)**
2. If $\gamma = 0.6$, $\sigma x = 1.5$ and $\sigma y = 2$, Find the bxy and byx. **(Unit-I, Prob.8)**
3. Define Index Numbers. **(Unit-II, SQA.3)**
4. From the following data construct an Index number for 1994 taking 1999 as base as per simple Aggregative method.

Year	Commodity					
	A	B	C	D	E	F
1970	45	60	20	50	85	120
1975	55	70	30	75	90	130

(Unit-II, Prob.2)

5. What are the characteristics of time series ? **(Unit-III, SQA.2)**
6. The following table gives the annual sales (in Rs.'000) of a commodity :

Year	Sales
1990	710
1991	705
1992	680
1993	687
1994	757
1995	629
1996	644
1997	783
1998	781
1999	805
2000	805

Determine the trend by calculating the 5-yearly moving average. **(Unit-III, Prob.11)**

7. Mutually Exclusive Events (Unit-IV, SQA.4)

8. If the probability of a defective bot is 0.2, find (i) mean (ii) standard deviation for the distribution of bolts in a total of 400. (Unit-I, Prob.6)

PART - B (5 × 12 = 60 Marks)

Note: Answer all the questions in not exceeding 4 pages each by using internal choice.

9. (a) Define the principle of least squares and standard error of estimate. (Unit-I, Q.No.9)

OR

(b) The correlation co-efficient between x and y is (r) = 0.6, σ_x = 1.5 σ_y = 2, $\overline{X} = 10$ and $\overline{Y} = 20$, find the two Regression Equations. (Unit-I, Prob.12)

10. (a) Construct index numbers of price from the following data by applying :

(i) Laspeyres method

(ii) Paasche method

(iii) Fisher's ideal method, and

(iv) Marshall-Edgeworth method

Commodity	2010		2011	
	Price	Quantity	Price	Quantity
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

(Unit-II, Prob.10)

OR

(b) The following are the indices (2007. Base)

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Indices	100	120	122	116	120	120	137	136	149	156	137

Shift the base to 2012 and recast the index numbers. (Unit-II, Prob.18)

11. (a) What are the Components of Time Series? (Unit-III, Q.No.4)

OR

(b) Fit a straight line by the Least Square Method and tabulate the trend values for the above data.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	77	88	94	85	91	98	90

(Unit-III, Prob.18)

12. (a) There are 12 cards numbered 1 to 12 in a box, if two cards are selected what is the probability that sum is odd

(a) With replacement ?

(b) Without replacement ? **(Unit-IV, Prob.25)**

OR

(b) A fair coin is tossed 4 times. Find the probability that there will appear.

a) 2 heads

b) 1 tail and 3 heads

c) at least one head

d) exactly one head

e) Not more than one head **(Unit-IV, Prob.12)**

13. (a) The following table gives the number of days in a 50 day period during which automobile accidents occurred in a certain part of a city. Fit a Poisson distribution to the data.

No. of accidents	0	1	2	3	4
No. of days	19	18	8	4	1

(Unit-V, Prob.17)

OR

(b) A study of past participants indicates that the mean length of times spent on the programme is 500 hours; and that, this normal distribution random variable has a standard deviation of 100 hours. What is the probability that a participant selected at random will required to complete the programme in following cases

(i) 'More' than 500 hrs.

(ii) Between 500 and 650 hours.

(iii) Between 550 and 650 hours.

(iv) Less than 580 hours

(v) Between 420 and 570 hours **(Unit-V, Prob.20)**

FACULTIES OF COMMERCE

B.Com. IV – Semester (CBCS) Examination

(Common Paper for General / Computers / Computer Applications / Advertising / Foreign Trade / and Tax Procedure Courses)

MODEL PAPER - II

BUSINESS STATISTICS - II

Time : 3 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

Note: Answer any five of the following questions not exceeding 20 lines each.

ANSWERS

1. What are the properties of regression coefficient? **(Unit-I, SQA.7)**
2. Given the two regression coefficients $b_{yx} = 0.4$ and $b_{yx} = 0.9$, find the value of correlation Coefficient. **(Unit-I, Prob.1)**
3. Explain the importance of index numbers. **(Unit-II, SQA.1)**
4. From the following data calculate a price Index based on price Relatives Method using Arithmetic Mean.

Commodity	A	B	C	D	E	F
Price 2015(Rs.)	45	60	20	50	85	120
Price 2016 (Rs.)	55	70	30	75	90	130

(Unit-II, SQA.1)

5. What is time sereis? **(Unit-III, SQA.1)**
6. Calculate three year moving average for the following data:

Year :	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
Value :	242	250	252	249	253	255	251	257	260	265	262

(Unit-III, Prob.12)

7. Random Experiment. **(Unit-IV, SQA.7)**
8. Comment on the following:

   For a Binomial Distribution Mean = 7 and Variance = 11. **(Unit-V, Prob.5)**

PART - B (5 × 12 = 60 Marks)

Note: Answer all the questions in not exceeding 4 pages each by using internal choice.

9. (a) Differentiate between linear and non-linear regression. **(Unit-I, Q.No.7)**

OR

(b) Given the two regression coefficients $b_{yx} = 0.4$ and $b_{yx} = 0.9$, find the value of correlation Coefficient. **(Unit-I, Prob.9)**

10. (a) From the following data calculate price index according to

(i) Laspeyre,

(ii) Paasche and

(iii) Marshall-Edgeworth methods.

Item	Base year		Current year	
	Price (`)	Expenditure (`)	Price (`)	Expenditure (`)
A	5	50	8	40
B	7	25	12	30
C	9	10	15	25
D	12	5	20	18

(Unit-II, Prob.11)

OR

(b) The index of 2010 is 100. It rises by 5% in 2011, falls 2% in 2012, rises 4% in 2013, rises 10% in 2014, falls 3% in 2015, and rises 8% in 2016. Find out the indices for the 7 years assuming that all the increases and decreases are the percentages of the respective proceeding years - Also recast the indices shifting base to 2014. **(Unit-II, Prob.17)**

11. (a) What are the various methods of time series? Explain in detail semi average method. With an example. **(Unit-III, Q.No.9)**

OR

(b) Production figure of a Textile Industry are as follows.

Year	2011	2012	2013	2014	2015	2016	2017
Production (in '000 units)	12	10	14	11	13	15	16

For the above data

(i) Determine the straight line equation under the Least Square Method.

(ii) Find the Trend Values and show the trend line on a graph paper. **(Unit-III, Prob.19)**

12. (a) A bag A contains 2 white and 3 red balls and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the red ball drawn is from bag. **(Unit-IV, Prob.29)**

OR

(b) From 30 tickets marked with first 30 numerals, 1 ticket is drawn at random. It is then replaced and a second draw is made. Find the probability that in the first draw it is multiple of 5 or 7 and in second draw it is a multiple of 3 or 7. **(Unit-IV, Prob.27)**

13. (a) The number of defects per unit in a sample of 330 units of manufactured product was found as follows:

No. of defects :	0	1	2	3	4
No. of units :	214	92	20	3	1

(Unit-V, Prob.16)

OR

(b) What are the features and assumptions of Poisson distribution. **(Unit-V, Q.No.7)**

FACULTIES OF COMMERCE

B.Com. IV – Semester (CBCS) Examination

(Common Paper for General / Computers / Computer Applications / Advertising / Foreign Trade / and Tax Procedure Courses)

MODEL PAPER - III

BUSINESS STATISTICS - II

Time : 3 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

Note: Answer any five of the following questions not exceeding 20 lines each.

ANSWERS

1. What is the Limitations of Regression Analysis? **(Unit-I, SQA.8)**
2. What is Marshall Edgeworth method ? **(Unit-II, SQA.7)**
3. Reconstruct the series of index numbers given below by shifting the base to 2010.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
Index No	100	120	132	140	150	164	180	208	220

(Unit-II, Prob.16)

4. Explain the utility of time series. **(Unit-III, SQA.12)**
5. From the following data, calculate trend values using Four Yearly Moving Averages.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Production (in tons)	506	620	1036	673	588	696	1116	738	663

(Unit-III, Prob.15)

6. What is the probability for a leap year to have 52 Mondays and 53 Sundays? **(Unit-IV, Prob.1)**
7. Explain the importance of probability. **(Unit-IV, SQA.8)**
8. Importance of Binomial Distribution. **(Unit-V, SQA.10)**

PART - B (5 × 12 = 60 Marks)

Note: Answer all the questions in not exceeding 4 pages each by using internal choice.

9. (a) Differentiate between Correlation and Regression. **(Unit-I, Q.No.8)**

OR

(b) 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

(Unit-I, Prob.11)

10. (a) What are the problems involved in construction of index numbers ? Explain. **(Unit-II, Q.No.5)**

OR

(b) Compute Price Index number by using:

(i) Paasches and

(ii) Marshal and Edge worth methods.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
P	5	100	6	150
Q	4	80	5	100
R	2	60	5	72
S	12	30	9	33

(Unit-II, Prob.12)

11. (a) Define least square method. Explain merits and demerits of least square method. **(Unit-III, Q.No.12)**

OR

(b) Calculate the 3 yearly the 5 yearly moving averages for the following time series.

Year :	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Production (in quintals)	500	540	550	530	520	560	600	640	620	610	640

(Unit-III, Prob.16)

12. (a) A factory has two machines. Empirical evidence has established that machines I and II produce 30% and 70% of output respectively. It has also been established that 5% and 1% of the output produced by these machines respectively was defective. A defective item is drawn at random. What is the probability that the defective item was produced by either machine I or machine II ? **(Unit-IV, Prob.30)**

OR

(b) A bag contains 4 defective and 6 good Electronic Calculators. Two calculators are drawn at random one after the other without replacement. Find the probability that

i) Two are good

ii) Two are defective and

iii) One is good and one is defective. **(Unit-IV, Prob.19)**

13. (a) Explain in detail about Central Limit Theorem. **(Unit-V, Q.No.16)**

OR

(b) Ten unbiased coins are tossed simultaneously. Find the probability of obtaining:

i) Exactly 6 Heads

ii) Atleast 8 Heads

iii) No Heads

iv) Atleast one Head

v) Not more than 3 Heads and

vi) Atleast 4 heads. **(Unit-V, Prob.10)**