SIA®
Publishers & Distributors Pvt. Ltd.
SPECTRUM ALL-IN-ONE® Series

REVISED 2022 EDITION

BUSINESS STATISTICS-II

B.Com II-Year IV-Sem.
(Common to All Courses (DSC-402))
As Per the Latest (2019-20) Syllabus (CBCS)

B.Com : II-Year IV-Sem.

Exam QP's : **Latest QPs :** Sept./Oct.-21 (OU) | July/Aug.-21 (KU) | July/Aug.-21 (MGU) | Jan.-21 (OU)
Previous QP's : May/June-18 (KU), May/June-19 (MGU), May/June-18 (OU), May/June-19 (OU)

Prepared by : SIA Team of Experts

Total Pages : 192

Salient Features:

- Complete Syllabus covered as per the Latest Regulation of Osmania University.
- Subjects are Explained in Simple and Easy to Understand Language.
- Exhaustive coverage of Topics from Examination Point of View.
- Gives Complete Knowledge of the Syllabus Topics.
- Every unit is structured into two main sections viz., Short Questions (Part-A) and Essay Questions (Part-B) and Answers.
- Internal Assessment (Internal or Exam) Patterns are included.
- List of Important Definitions and Formulae.
- Unit-wise Frequently Asked Questions and Important Questions.
- Previous to Latest Exams Question Papers and Answers (KU, MGU, OU).

Syllabus Covered for :

- Osmania University
- Kakatiya University
- Mahatma Gandhi University
- Palamuru University
- Telangana University
- Satavahana University

Books Available for B.Com II-year IV-Sem (O.U)

- English (First Language)
- Hindi/Telugu/Sanskrit (Second Language)

Skill Enhancement Course (SEC 3)

(a) Leadership & Management Skills

Skill Enhancement Course (SEC 4)

(a) Practice of Life and General Insurance

Discipline Specific Course (DSC)

- Income Tax (DSC-401)
- Business Statistics-II (DSC-402)
- Corporate Accounting (DSC-403) (General/Foreign Trade/Honours)
- Web Technologies (DSC-403) (Com. App.)

Printed at:
SIA Publishers & Distributors Pvt. Ltd.,
Shed No. C-8/2, IDA Uppal, Hyderabad-39.

SIA Publishers & Distributors Pvt. Ltd.
Office Address:
#5-1-911/5, Biyabani Square Building, Putli Bowli,
Lane Opp. to Osmania Medical College, Koti,
Hyderabad - 500 095, Telangana, India

PREFACE

ABOUT THE SUBJECT

The subject **Business Statistics-II** includes all those topics which help to analyze the numerical information and figures on scientific basis. Statistics is defined in plural sense which means set of numerical data or figure. However, it carries out various functions and plays an important role in several areas like, economies, national income accounting, planning, business etc. Thus, statistics is used to gather, represent and examine the numerical figures on scientific basis.

The main objective behind introducing the subject **Business Statistics-II** in B.Com course is to make students acquire both theoretical and practical knowledge of the statistical topics.

The important topics discussed in this subject are as follows,

- Concept of Regression.
- Concept of Index Numbers.
- Concept of Time Series.
- Concept of Probability with Approaches and Theorems.
- Concept of Binomial, Poisson and Normal Distribution.

ABOUT THE BOOK

The book entitled **'Business Statistics-II'** is designed for B.Com II-Year IV-Semester students. The content provided in this book is strictly as per the latest syllabus prescribed by Osmania University.

Every concept is explained in a simple manner with sufficient number of examples so as to facilitate better understanding and easy learning in a shorter span of time. Keeping in view the examination pattern of B.Com students, this book provides the following features,

- Exclusively Prepared as per the **Latest (2019-20) Syllabus (CBCS)** Prescribed by the University.
- Unit-wise **List of Definitions and Formulae** are listed separately.
- Every unit is structured into **two main sections** viz., **Short Questions (Part-A)** and **Essay Questions (Part-B)** answers.
- **Exercise and Practice Questions** are Listed at the end of every Unit.
- Unit-wise **Internal Assessment** (Internal Exam) Pattern is attached with every unit.
- **Unit-wise Frequently Asked Questions (FAQs) and Important Questions (IQs)** are listed at the end.

An attempt has been made through this book to present theoretical and practical knowledge of "**Business Statistics-II**". This book is especially prepared for undergraduate students.

The table below illustrates the complete idea about the subject, which will be helpful to plan and score good marks in the end examinations.

S.No.	Unit Name	Description
1.	Regression	This unit covers the topics: 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.
2.	Index Numbers	This unit covers the topics: 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.
3.	Time Series	This unit covers the topics: Introduction - Components – Methods-Semi Averages - Moving Averages – Least Square Method - Deseasonalisation of Data – Uses and Limitations of Time Series.
4.	Probability	This unit covers the topics: 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.
5.	Theoretical Distributions	This unit covers the topics: 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).

It is sincerely hoped that this book will satisfy the expectations of students and at the same time helps them to score maximum marks in exams.

Suggestions for improvement of the book from our esteemed readers will be highly appreciated and incorporated in our forthcoming editions.

BUSINESS STATISTICS-II

B.Com II-Year IV-Semester

CONTENTS

Syllabus

(Subject: Business Statistics-II, B.COM IV Semester, (DSC-402) (2019-20))

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

THEORETICAL 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).

LIST OF IMPORTANT DEFINITIONS AND FORMULAE

UNIT - I

1. According to M.M. Blair, "Regression is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data".
2. Linear regression is a form of regression which is used for modeling the relationship between scalar variables like 'X' and 'Y'.
3. In the non-linear regression the explained variable (dependent variable) changes at varying rate with a given change in the explaining variable (independent variable).
4. The regression coefficient X on Y measures the change in X corresponding to a unit change in Y and the regression coefficient of Y on X measures the change in Y corresponding to a unit change in X,

   $r.\frac{\sigma_y}{\sigma_x}$ *and* $r.\frac{\sigma_x}{\sigma_y}$ are known as coefficient of regression.

5. "$r.\frac{\sigma_y}{\sigma_x}$" is denoted by "$b_{xy}$" and is given by,

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

UNIT - II

1. According to Maslow, "Index number is a numerical value characterizing the change in complex economic phenomena over a period of time or space".
2. Single Price Index = $\frac{\text{Current value}}{\text{Base value}} \times 100$ or $\frac{P_1}{P_0} \times 100$
3. In a weighted aggregate price index, certain weight is assigned to each and every commodity or item of group in accordance with its significance.
4. The formula for Laspeyre's price index method is as follows,

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

5. The formula for Paasche's index method is,

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

6. A statistician named "Fisher" introduced this method, which is a geometric mean of Laspeyre's and Paasche's methods. The formula used for this method is,

$$P_{01} + \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

UNIT - III

1. Ya-Lun-Chou has defined "time series as a collection of readings belonging to different time periods, of some economic variable or composite of variables".
2. Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3} \ldots$$

3. Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5} \ldots$$

4. Least square method is a statistical procedure which is used to find the best fit curve for the set of data where different variables are involved.

5. Deseasonalisation of data refers to the process of eliminating seasonal fluctuations from the given time series for, it facilitates in adjusting the given time series for seasonal fluctuations and therefore left out with variables like trend component, cyclical and irregular variations.

UNIT - IV

1. Probability can be defined as the chance or 'likelihood of occurrence' of an experiment or event. Probability of any event ranges from 0 to 1.

2. Joint probability of A and B is represented as $P(A \cap B)$

$\therefore \quad P(A \cap B) = P(A).P(B)$

3. The number of ways of selecting some objects (r) from total number of distinguishable objects (D) which can be arranged in an order is called 'permutation'. It is denoted as '${}^nP_r$'.

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

4. The number of ways of selecting 'r' objects from 'n' different objects irrespective of their arrangement is called as 'combinations'. It is denoted as ${}^nC_r$.

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

5. An event is a possible outcome of an experiment or a result of trial. Basically there are two types of events. Simple and compound event.

UNIT - V

1. Binomial distribution is a discrete probability distribution developed by a Swiss mathematician, 'James Bernoulli' in 1700. It is used for finite or limited number of trails 'n'. It produces successes and failures based on two parameters 'n' and 'p'.

2. The mean of binomial distribution,

$$\mu = np$$

3. Standard deviation is given by,

$$\sigma = \sqrt{npq}$$

4. 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., $p \to o$)

5. Always the sum of infinite probabilites in poisson distribution is 1 i.e.,

$$P(0)+P(1)+P(2)+\ldots+P(\infty)=1$$

6. The probability of 'X' occurrences in poisson distribution is given by,

$$P(X) = \frac{e^{-\lambda}.\lambda^x}{x!}$$

UNIT 1

REGRESSION

SYLLABUS

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.

LEARNING OBJECTIVES

- ✓ Concept, Types and Applications of Regression.
- ✓ Linear and Non-Linear Regression.
- ✓ Differences Between Correlation and Regression.
- ✓ Lines of Regression.
- ✓ Derivation of Line of Regression of Y on X.
- ✓ Derivation of Line of Regression of X on Y.
- ✓ Use of Regression Lines for Prediction.

INTRODUCTION

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.

Linear regression is a form of regression which is used for modelling the relationship between scalar variables like 'X' and 'Y'. 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.

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 modelled by a function i.e., a non-linear blend of model parameters and depends on one or more independent variable.

In order to predict the value of variable 'Y' (dependent variable) for a given value of variable 'X' (independent variable), the equation of regression line is preferable.

PART-A

SHORT QUESTIONS AND ANSWERS

Q1. Define Regression Analysis.

Answer :

According to Ya-Lun 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".

According to Morris Hamburg, "the term 'regression analysis' refers to the methods by which estimates are made of the values of 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".

Based on the above two definitions, regression analysis can be defined as a measure of the average relationship between two or more variables in terms of original units of the data.

Q2. Features of Regression Coefficients.

Answer : May/June-19, Q1(MGU)

The properties/features of regression coefficients are as follows,

1. **Deviations**

The deviations are taken from the two means, i.e., Actual mean and Assumed mean.

When deviations taken from actual mean,

$r\frac{\sigma_x}{\sigma_y}$ is denoted by "b_{xy}" and is given by,

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

$r\frac{\sigma_y}{\sigma_x}$ is denoted by "b_{yx}" and is given by,

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

When deviations taken from assumed mean,

$$b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$$

$$b_{xy} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dy^2 - (\Sigma dy)^2}$$

2. **Least Square Line**

The least square regression line always passes through $(\bar{x}, \bar{y})$.

3. **Same Sign**

The regression coefficients have the same signs.

4. **Sign of 'r'**

The correlation coefficient has the same sign as that of regression coefficients.

5. **Independent of origin**

The regression coefficients are independent of change of origin but not of scale.

6. **r is Geometric Mean**

The correlation coefficient (r) is the geometric mean of two regression coefficients.

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

$$r = \sqrt{r.\frac{\sigma_y}{\sigma_x} \times \frac{\sigma_x}{\sigma_y}}$$

Q3. State the differences between Correlation and Regression.

Answer : Sept/Oct-21, Q1(OU)

The following are the differences between correlation and regression,

S.No	Basis	Correlation	Regression
1.	Meaning	It refers to the relationship between two or more variables which vary in sympathy so that the movements in one variable tend to be accompanied by the corresponding movements in others.	It refers to stepping back or returning to average value and is a mathematical measure expressing the average relationship between the variables.
2.	Nature	It is a measure of the 'degree and direction' of relationship between the variables.	It studies 'nature' of relationship between the variables.
3.	Cause and Effect Relationship	It does not indicate the cause and effect relationship between the variables.	It clearly indicates the cause and effect relationship between the variables.
4.	Variables	It does not describe which variable is the dependent variable and which is the independent variable.	The variable corresponding to cause is taken as independent variable and the variable corresponding to effect is taken as dependent variable.
5.	Relative and Absolute Measures	Its coefficient are relative measures of the linear relationship.	Its coefficients are absolute measures indicating the change in the value of one variable for a unit change in the value of the other variable.

Q4. Write about principle of least squares and standard error of estimate.

Answer :

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. Lines of regression uses the principle of least squares to give the best fit line for estimating the value of one variable given the value of another variable.

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 perfect. A measure of precision of these estimate is given by the standard error of the estimate. Standard deviation gives us measure of dispersion of the observations about the mean of the distribution whereas standard error of estimate gives us a measure of the observations about the line of regression.

Formula of Standard Error of Estimate is as follows,

$$\text{Standard error of estimate } (S_{yx}) = \sqrt{\frac{\Sigma(Y - Y_c)^2}{N}}$$

Q5. Write three limitations of regression analysis.

Answer : May/June-18, Q1(a) (KU)

Some of the limitations of regression analysis are as follows,

1. Social Sciences

Regression analysis assumes that linear relationship exists among the related variables. But in the area of social sciences, linear relationship may not exist among the related variables.

2. Evaluation of Dependent Variable

When regression analysis is used to evaluate the value of dependent variable based on independent variable, it is assumed that the static conditions of relationship exist between them. These statistic conditions do not exist in social sciences, so this assumption minimizes the use of regression analysis in social science.

3. Inaccurate Results

The value of dependent variable can be evaluated based on independent variable by using regression analysis but only upto some limits. If the circumstances go beyond the limits, then results would be inaccurate.

Q6. Given the two Regression Co-efficients b_{yx} = 0.4 and b_{xy} = 0.9, find the value of Correlation Co-efficient (r).

Answer : Sept./Oct.-21, Q6 (OU)

Given, $b_{yx} = 0.4,$

$b_{xy} = 0.9$

We know that, correction co-efficient (r),

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

$$r = \sqrt{0.9 \times 0.4}$$

$$r = \sqrt{0.36} = 0.6$$

∴ Correlation Co-efficient (r) = 0.6

Q7. If x = 0.85y and y = 0.89x Find the coefficient of correlation.

Answer : Jan.-21, Q1 (OU)

Given,

$x = 0.85y$ or $bxy = 0.85$

$y = 0.89x$ or $byx = 0.89$

Coefficient of Correlation $(r) = \sqrt{bxy \times byx}$

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

$$= \sqrt{0.7565}$$

$$r = 0.869 \text{ or } 0.87.$$

Q8. If r = 0.8; σ_x = 2.5, σ_y = 3.5, find b_{xy} and b_{yx}.

Answer : May/June-19, Q1 (OU)

Given that,

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

(i) $$b_{xy} = r \times \frac{\sigma_x}{\sigma_y}$$

$$= 0.8\left(\frac{2.5}{3.5}\right)$$

$$= 0.8 \times 0.714$$

$$= 0.571$$

(ii) $$b_{yx} = r \times \frac{\sigma_y}{\sigma_x}$$

$$= 0.8\left(\frac{3.5}{2.5}\right)$$

$$= 0.8 \times 1.4$$

$$= 1.12$$

Q9. If γ = 0.6, σ_x = 1.5 and σ_y = 2, find the b_{xy} and b_{yx}.

Answer : May/June-18, Q1 (OU)

Given that,

$r = 0.6, \sigma_x = 1.5$ and $\sigma_y = 2$

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

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

$$= 0.45$$

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

$$b_{yx} = 0.6 \times \frac{2}{1.5}$$

$$= 0.8$$

Q10. Co-efficient of correlation = 0.60, σ_x = 1.5, σ_y = 2.0, x = 10, y = 20, find regression equation y on x.

Answer : May/June-18, Q1(g) (KU)

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

$$b_{yx} = 0.60 \times \frac{2}{1.5}$$

$$b_{yx} = 0.8$$

Regression Equation of y on x

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

$$y - 20 = 0.8(x - 10)$$

$$y - 20 = 0.8x - 8$$

$$y = 0.8x - 8 + 20$$

$$y = 0.8x + 12$$

PART-B

ESSAY QUESTIONS AND ANSWERS

1.1 REGRESSION - INTRODUCTION

Q11. Define and explain regression analysis? Explain regression variables and types of regression.

Answer :

Regression Analysis

According to Ya-Lun 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".

According to Morris Hamburg, "the term 'regression analysis' refers to the methods by which estimates are made of the values of 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".

Based on the above two definitions, regression analysis can be defined as a measure of the average relationship between two or more variables in terms of original units of the data.

Regression analysis is a technique which expresses the relationship between two or more variables in equation form to estimate the value of variable based on the given value of another variable.

Regression takes its name from studies made by Sir Francis Galton. He compared the heights of persons to the heights of their parents. His major conclusion was that the children of unusually tall persons tend to be shorter than their parents while children of short parents tend to be shorter than their parents.

In a sense, the successive generations of off-springs from tall persons "regress" downward toward the height of the population. While the reverse is true originally for short families. But the distribution of heights for the total population, continues to have the same variability from generation to generation.

Regression Variables

The regression variables are as follows,

1. **Independent Variable (Regressor or Predictor or Explanatory)**

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.

2. **Dependent Variable (Regressed or Explained 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.

Types of Regression

The various types of regression are as follows,

1. **Simple Regression**

The regression analysis confined to the study of only two variables at a time is termed as simple regression.

2. **Multiple Regression**

The regression analysis for studying more than two variables at a time is termed as multiple regression.

3. **Linear Regression**

The regression curve which has a straight line, then such a regression is termed as linear regression. The equation of such a curve is the equation of a straight line i.e., first degree equation in variables x and y.

4. **Nonlinear Regression**

The regression curve which does not has a straight line, then such as regression is termed as curved or non-linear regression. The regression equation will be a functional relation between variables x and y involving terms in x and y of degree more than one.

Q12. What is meant by regression? What is the importance and limitations of Analysis? Jan-21, Q9 (OU)

OR

Define Regression. Discuss the uses and limitations of regression analysis.

Answer : July/Aug-21, Q2 (KU)

Regression Analysis

Regression analysis is a technique which expresses the relationship between two or more variables in equation form to estimate the value of variable based on the given value of another variable.

According to M.M. Blair, "Regression is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data".

Importance/Applications/Uses of Regression Test

Regression lines or equations are useful in the predictions of values of one variable for a specified value of the other variable. Some of the applications/ utility of regression test are as follows,

1. **Pharmaceutical Firms**

   It helps pharmaceutical firms which are interested in studying the effect of new drugs in patients.

2. **Predict the Future Demand**

   It helps to estimate or predict the future demand for a specified price when price and demand are inter related.

3. **Crop Yield**

   It helps to predict crop yield for a particular amount of rainfall when crop yield depends on the amount of rainfall.

4. **Advertising Expenditure**

   It helps in estimating the advertising expenditure for a required amount of sales or sales expected for a particular advertising expenditure.

5. **Capital and Profit**

   It helps to predict profits for a specified amount of capital invested when capital employed and profits earned.

Limitations of Regression Analysis

Some of the limitations of regression analysis are as follows,

1. **Social Sciences**

   Regression analysis assumes that linear relationship exists among the related variables. But in the area of social sciences, linear relationship may not exist among the related variables.

2. **Valuation of Dependent Variable**

   When regression analysis is used to evaluate the value of dependent variable based on independent variable, it is assumed that the static conditions of relationship exist between them. These statistic conditions do not exist in social sciences, so this assumption minimizes the use of regression analysis in social science.

3. **Inaccurate Results**

   The value of dependent variable can be evaluated based on independent variable by using regression analysis but only upto some limits. If the circumstances go beyond the limits, then results would be inaccurate.

Q13. Write the relation between correlation and regression.

Answer : May/June-18, Q2(a) (KU)

The following points highlights the relation between correlation and regression.

1. They are key statistical tools for studying the functional relationship between two or more variables. They both help in determining the nature and strength of relationship between the variables.
2. They helps the decision makers in prediction and reduction of uncertainty.
3. They are interdependent.
4. They can be demonstrated with the help of graphs which are known as scatterplots.
5. They could be either positive or negative.
6. They both involve straight line relationship.
7. They represent continuous scores on both variables.
8. They helps to measures the errors in a similar way.
9. They are linked together i.e, The correlation coefficient 'r' is linked to the coefficient of determination 'R^2' in the regression analysis.
10. They use same signs. The correlation coefficient 'r' takes same sign as taken by 'b' in regression analysis.

1.1.1 Linear and Non-Linear Regression

Q14. What do you mean by linear and non-linear regression? Distinguish between them.

Answer :

Linear Regression

Linear Regression is a form of regression which is used for modeling the relationship between scalar variables like '*X*' and '*Y*'. Under linear regression, linear functions are used to model the data and the unknown parameters of models are estimated from the data. Therefore, these models are known as Linear Models.

Linear Models commonly refers to those models, where the conditional mean of variable '*Y*' 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 quartile of the conditional distribution of '*Y*' 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 '*Y*' for a given '*X*', instead of joint probability distribution of '*Y*' and '*X*'.

Non-Linear Regression

Non-Linear Regression refers to the explained variable (dependent variable) which 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'.

Differences between Linear and Non-linear Regression

The differences between Linear and Non-Linear Regression are as follows,

S.No.	Basis	Linear Regression	Non-Linear Regression
1.	Meaning	The regression which is used for modelling the relationship between a scalar variable 'X' and 'Y' is referred as linear regression.	The regression where the observational data are modeled by a function i.e., a nonlinear blend of model parameters is referred as Non-linear regression.
2.	Curve	The regression curve which has a straight line, then such a regression is termed as linear regression.	The regression curve which does not has a straight line, then such a regression is termed as curved or nonlinear regression.
3.	Model form	The parameters are considered as linear combinations.	The parameter are considered as functions.
4.	Solution	The solution for parameters are represented as closed form.	The parameters needs 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	The estimation of parameters are unbiased in case of un-correlated errors.	The estimation of parameters are usually biased incase of un-correlated errors.
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 of degree more than one.

1.2 CORRELATION VS REGRESSION

Q15. Define regression and what are the differences between correlation and regression.

May/June-18, Q9(a) (OU)

OR

Distinguish between regression and correlation.

(Refer Only Topic: Differences between correlation and Regression)

Answer :

July/Aug-21, Q1(a) (KU)

Regression

According to M.M. Blair, "Regression is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data".

Differences between Correlation and Regression

The differences between Correlation and Regression are as follows,

S.No	Basis	Correlation	Regression
1.	Meaning	It refers to the relationship between two or more variables which vary in sympathy so that the movements in one variable tend to be accompanied by the corresponding movements in others.	It refers to stepping back or returning to average value and is a mathematical measure expressing the average relationship between the variables.
2.	Nature	It is a measure of the 'degree and direction' of relationship between the variables.	It studies 'nature' of relationship between the variables.
3.	Cause and Effect Relationship	It does not indicate the cause and effect relationship between the variables.	It clearly indicates the cause and effect relationship between the variables.
4.	Variables	It does not describe which variable is the dependent variable and which is the independent variable.	The variable corresponding to cause is taken as independent variable and the variable corresponding to effect is taken as dependent variable.
5.	Relative and Absolute Measures	Its coefficient are relative measures of the linear relationship.	Its coefficients are absolute measures indicating the change in the value of one variable for a unit change in the value of the other variable.
6.	Prediction	It cannot be used for predicting or estimating value.	It is very helpful in predicting and estimating value of one variable given the value of another variable.
7.	Coefficients	Its coefficients are symmetric i.e., $r_{yx} = r_{xy}$.	Its coefficients are asymmetric i.e., $b_{xy} \neq b_{yx}$.
8.	Relationship	Its coefficient can be calculated from regression coefficients.	Its coefficients cannot be directly compared from correlation coefficient.
9.	Range/Limit	The range of r is +1 to – 1.	The range of b_{xy} and b_{yx} is not restricted.
10.	Drawbacks	There may be inaccurate correlation between variables due to chance.	There is no such thing in regression.

1.3 LINES OF REGRESSION - DERIVATION OF LINE OF REGRESSION OF Y ON X - LINE OF REGRESSION OF X ON Y

Q16. What do you mean by lines of regression? Derive the equation of lines of regression.

Answer :

Lines of Regression

"The regression lines show the average relationship between two variables". *–Galton*

"The device used for estimating the value of one variable from the value of the other consists of line through the points, drawn in such a manner as to represent the average relationship between the two variables. Such a line is called the line of regression".

–J.R. Stockton

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 as 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 Curvi-Linear Regression.

The lines of regression gives the best estimate to the value of a variable for any given value of another variable. Thus it is the line of "best fit" which is obtained by using the principles of least squares.

Derivation of Lines of Regression of Y on X

Let X and Y be two variables. Also assume that Y is dependent variable and X is independent variable and the bi-variate distribution is (x_i, y_i) for $i = 1, 2, ... n$.

Let the line of regression of Y on X be,

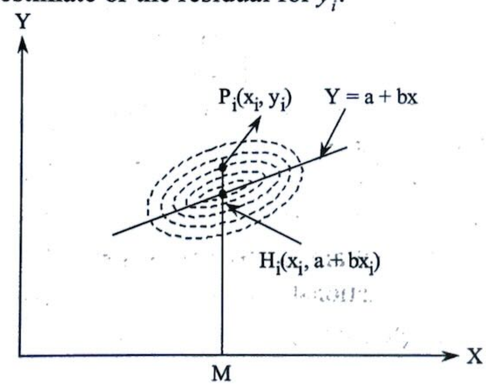
$$Y = a + bX \quad ... (1)$$

Where a and b are constants. We need to find the values of a and b such that this line is the line of "best fit".

Let $P_i(x_i, y_i)$ be any point in the scatter diagram. Now draw a line P_iM perpendicular to the x-axis so that it meets the line in equation (1). The intersecting point of the two lines is H_i whose coordinates are $(x_i, a + bx_i)$.

Thus, $P_iH_i = P_iM - H_iM = y_i - (a + bx_i)$

This is called the error of estimate or the residual for y_i.



According to the principle of least squares we have to determine a and b such that E is minimum.

$$E = \sum_{i=1}^{n} P_iH_i^2 = \sum_{i=1}^{n}(y_i - a - bx_i)^2$$

The partial derivative of E with respect to a and b is,

$$\frac{\partial E}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i)$$

$$\Rightarrow \quad \frac{0}{2} = -\left[\left(\sum_{i=1}^{n} y_i\right) - na - b\sum_{i=1}^{n} x_i\right]$$

$$\Rightarrow \quad 0 = -\sum_{i=1}^{n} y_i + na + b\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \quad ... (2)$$

$$\Rightarrow \quad \frac{\partial E}{\partial b} = -2\sum_{i=1}^{n} x_i(y_i - a - bx_i)$$

$$\Rightarrow \quad \frac{0}{2} = -\left[\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2\right]$$

$$\Rightarrow \quad 0 = -\sum_{i=1}^{n} x_i y_i + a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \quad \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{... (3)}$$

The equations (2) and (3) are known as the normal equations for estimating a and b.

Divide equation (2) by n, we get,

$$\bar{y} = a + b\bar{x} \quad \text{... (4)}$$

This equation indicates that the regression line of Y on X passes through the point $(\bar{x}, \bar{y})$.

Now,

$$\mu_{11} = \text{Cov}(X, Y)$$

$$\mu_{11} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}$$

$$\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \mu_{11} + \bar{x}\,\bar{y} \quad \text{... (5)}$$

And $\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$

$$\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sigma_x^2 + \bar{x}^2$$

Divide equation (3) by n, we get,

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i = a.\frac{1}{n}\sum_{i=1}^{n} x_i + b.\frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad \text{... (6)}$$

Equating equations (5) and (6), we get,

$$\mu_{11} + \bar{x}\,\bar{y} = a\bar{x} + b\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)\left[\because \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sigma_x^{\,2} + \bar{x}^2\right]$$

$$\mu_{11} + \bar{x}\,\bar{y} = a\bar{x} + b(\sigma_x^2 + \bar{x}^2) \quad \text{... (7)}$$

Multiply equation (4) by $\bar{x}$ and then subtract from equation (7), we get,

$$\Rightarrow \quad \mu_{11} + \bar{x}\,\bar{y} - \bar{x}\,\bar{y} = a\bar{x} + b(\sigma_x^2 + \bar{x}^2) - a\bar{x} - b\bar{x}^2$$

$$\Rightarrow \mu_{11} = b\sigma_x^2$$

$$\Rightarrow \quad b = \frac{\mu_{11}}{\sigma_x^2}$$

Since, the slope of the regression line of Y on X (i.e., equation (1)) is b and the line passes through the point $(\bar{x}, \bar{y})$ and, the equation for the line is,

$$Y - \bar{y} = b(X - \bar{x})$$

$$Y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2}(X - \bar{x})$$

$$Y - \bar{y} = r\frac{\sigma_Y}{\sigma_X}(X - \bar{x}) \quad \text{... (8)}$$

The equation (8) is the regression line of Y on X.

Derivation of Line of Regression Line of X on Y

Let the line of regression of X on Y is,

$$X = a + bY$$

The equation of the line of regression of X on Y can be obtained by following the same procedure used to obtain the equation of the line of regression of Y on X or by just interchanging the variables X and Y.

Thus, the equation of the line of regression of X on Y is,

$$X - \bar{x} = \frac{\mu_{11}}{\sigma_Y^2}(Y - \bar{y})$$

$$X - \bar{x} = r\frac{\sigma_X}{\sigma_Y}(Y - \bar{y})$$

1.3.1 Regression Coefficient

Q17. Explain briefly about Regression Coefficient.

Answer :

The regression coefficient X on Y measures the change in X corresponding to a unit change in Y and the regression coefficient of Y on X measures the change in Y corresponding to a unit change in X.

$r.\frac{\sigma_y}{\sigma_x}$ *and* $r.\frac{\sigma_x}{\sigma_y}$ are known as coefficient of regression.

The following are the four main methods to calculate regression coefficient X on Y (b_{xy}) or Y on X (b_{yx}),

1. Taking deviations from actual mean.
2. Taking deviations from assumed mean.
3. Applying actual observations or value based method.
4. Applying grouped data or step deviation method.

Let us discuss about each regression coefficient in detail,

1. Deviations from Actual Mean

The regression equation of X on Y is,

$$X - \bar{X} = r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$$

The regression Equation of Y on X is,

$$Y - \bar{Y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

$r\frac{\sigma_x}{\sigma_y}$ is called as regression coefficient of X on Y. It is denoted as b_{xy}.

$r\frac{\sigma_y}{\sigma_x}$ is calles as regression coefficient of Y on X. It is denoted as b_{xy}

2. **Deviations from Assumed Mean**

The deviations from assumed mean are taken when the mean solution results in fractions. The steps to calculate regression equations remains same but incase of assumed mean, the regression equations on X on Y is follows,

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

The value of r $\frac{\Sigma x}{\Sigma y}$ is as follows,

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

or

$$bxy = \frac{n\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{n\Sigma dy^2 - (\Sigma dy)^2}$$

Where,

$$dx = (X-a);\ dy = (Y-a)$$

3. **Applying Actual Observations**

The regression equations or lines can also be obtain through applying actual observations or value based method. The formula to apply actual observation of X on Y is as follows,

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

$$bxy = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}$$

4. **Applying Group Data**

The regression equations or lines can also be obtain through applying grouped data or step deviation method. The formula to apply grouped data of X on Y is as follows,

$$bxy = \frac{N.\Sigma fdxdy - \Sigma fdx\,\Sigma fdy}{N.\Sigma fd^2y - (\Sigma fdy)^2} \times \frac{ix}{iy}$$

Where,

i_x = Difference between class intervals of x

i_y = Difference between class intervals on y

Q18. Explain the properties of regression coefficient.

Answer :

The properties/features of regression coefficients are as follows,

1. **Deviations**

When deviations taken from actual mean,

"$r.\frac{\sigma_x}{\sigma_y}$ is denoted by "b_{xy}" and is given by,

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

"$r.\frac{\sigma_y}{\sigma_x}$ is denoted by "b_{yx}" and is given by,

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

When deviations taken from assumed mean,

$$b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$$

$$b_{xy} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dy^2 - (\Sigma dy)^2}$$

2. **Least Square Line**

The least square regression line always passes through $(\overline{x},\overline{y})$.

3. **Same Sign**

Both regression coefficients have the same sign.

4. **Sign of 'r'**

Correlation coefficient has the same sign as that of regression coefficients.

5. **Value**

If one value of regression coefficient is greater than unity, then the other must be less than unity, since r |>|1 (But both can be less than unity).

6. **Average Value**

Average value of the two regression coefficients will be greater than correlation coefficient.

7. **Independent of origin**

Regression coefficients are independent change origin but not for scale.

8. **r is Geometric Mean**

Correlation coefficient (r) is the geometric mean of two regression coefficients.

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

$$= \pm\sqrt{r.\frac{\sigma_y}{\sigma_x} \times r.\frac{\sigma_x}{\sigma_y}}$$

Q19. Write about the formulae's used in regression equations of X on Y an Y on X.

Answer :

The following are the formulae's used in regression equations of X on Y and Y on X,

S.No.	X on Y	Y on X
	Regression Equation	
1.	The regression equation which is generally used in regression analysis for X on Y is Xc = a + by.	The regression equation which is generally used in regression analysis for Y on X is Yc = a + bx.
	Arithmetic Mean	
2.	Arithmetic mean of X is, $\overline{X} = \frac{\Sigma x}{N}$	Arithmetic mean of Y is, $\overline{Y} = \frac{\Sigma y}{N}$
	Deviations from Actual Mean	
3.	The deviations if taken from actual mean, $X - \overline{X} = b_{xy}(Y - \overline{Y})$ $b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$	The deviations if taken from actual mean, $Y - \overline{Y} = b_{yx}(X - \overline{X})$ $b_{xy} = \frac{\Sigma xy}{\Sigma x^2}$
	Deviations from Assumed Mean	
4.	The deviations from assumed mean $(X - \overline{X}) = r\frac{\Sigma x}{\Sigma y}(Y - \overline{Y})$ the value of $r\frac{\Sigma x}{\Sigma y}$ or $b_{xy} = \frac{\Sigma dxdy - \frac{\Sigma dx \times \Sigma dy}{N}}{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}$ or $b_{xy} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dy^2 - (\Sigma dy)^2}$ Where, $dx = (X - a), dy = (Y - a)$	The deviations from assumed mean $(Y - \overline{Y}) = r\frac{\Sigma x}{\Sigma y}(X - \overline{X})$ the value of $r\frac{\Sigma x}{\Sigma y}$ or $b_{yx} = \frac{\Sigma dxdy - \frac{(\Sigma dx)(\Sigma dy)}{N}}{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}}$ or $b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2}$ Where, $dx = (X - a), dy = (Y - a)$
	Applying actual observations or value Based Method	
5.	The regression equation while applying actual observations, $X - \overline{X} = b_{xy}(Y - \overline{Y})$ $b_{xy} = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{N}}{\Sigma y^2 - \frac{(\Sigma y^2)}{N}}$ or $b_{xy} = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma y^2 - (\Sigma y)^2}$	The regression equation while applying actual observations, $Y - \overline{Y} = b_{yx}(X - \overline{X})$ $b_{yx} = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{N}}{\Sigma x^2 - \frac{(\Sigma x^2)}{N}}$ or $b_{yx} = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma x^2 - (\Sigma x)^2}$

Applying grouped data or step deviation method		
6.	The regression equation while applying grouped data, $b_{xy} = \frac{N.\Sigma fdxdy - \Sigma fdx.\Sigma fdy}{N.\Sigma fdx^2 (\Sigma fdx)^2} \times \frac{ix}{iy}$	The regression equation while applying grouped data, $b_{yx} = \frac{N.\Sigma fdxdy - \Sigma fdx.\Sigma fdy}{N.\Sigma fdx^2 (\Sigma fdx)^2} \times \frac{iy}{ix}$
Regression Coefficients		
7.	The deviations of regression coefficients, $b_{xy} = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(Y-\overline{Y})^2} = \frac{\Sigma xy}{\Sigma y^2}$	The deviations of regression coefficients, $b_{yx} = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2} = \frac{\Sigma xy}{\Sigma x^2}$
Correlation Coefficients		
8.	The calculation of correlation coefficient (r) $r = \sqrt{bxy \times byx}$	The calculation of correction coefficient (r) $r = \sqrt{byx \times bxy}$

1.4 USING REGRESSION LINES FOR PREDICTION

Q20. Explain briefly about the regression lines used for prediction.

Answer :

In order to predict the value of variable '*Y*' (dependent variable) for a given value of variable '*X*' (independent variable), the equation of regression line is preferable.

Example

The predicted value of variable '*Y*' (Y_i) for a given variable '*X*' (X_i) will be represented as,

$$Y_i = a + bX_i$$

Here, *a* and *b* are obtained from sample data and are least squares estimates derived from the normal equations.

Thus, it is important to use regression equations carefully for prediction and estimation. It is preferable to use regression equation for estimation when it properly fits data. Therefore, goodness is required to be tested before using lines of regression.

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. The 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. The linear regression is a good fit for the given data, If '*r*' is significant, then it is preferable to use line of regression for estimating '*Y*' for given '*X*'.
4. The 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.
5. It is useful to make predictions for linear regression model when sample data is drawn from the population.

PROBLEMS ON ACTUAL MEAN REGRESSION METHOD

Q21. From the following table obtain the two Regression Equations.

x	10	10	18	25	28	33	34	39	42	43
y	11	22	22	19	35	27	33	40	42	47

Solution :

Sept/Oct-21, Q9 (OU)

Calculation of Regression Equations

x	y	dx = x − $\bar{x}$ ($\bar{x}$ = 28)	dx²	dy = y − $\bar{y}$ ($\bar{y}$ = 30)	dy²	dxdy
10	11	−18	324	−19	361	342
10	22	−18	324	−8	64	144
18	22	−10	100	−8	64	80
25	19	−3	9	−11	121	33
28	35	0	0	5	25	0
33	27	5	25	−3	9	−15
34	33	6	36	3	9	18
39	40	11	121	10	100	110
42	42	14	196	12	144	168
43	47	15	225	17	289	255
Σx = 282	Σy = 298	Σdx = 2	Σdx² = 1360	Σdy = 2	Σdy² = 1186	Σdxdy = 1,135

$$\bar{x} = \Sigma\frac{x}{N} = \frac{282}{10} \cong 28 \ , \ \bar{y} = \frac{\Sigma y}{N} = \frac{298}{10} \cong 30$$

$$b_{xy} = \frac{\Sigma dxdy}{\Sigma dy^2} = \frac{1,135}{1186} = 0.96, \ b_{xy} = \frac{\Sigma dxdy}{\Sigma dx^2} = \frac{1,135}{1360} = 0.83$$

1. Regression Equation x on y

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

$$x - 28 = 0.96\,(y - 30)$$

$$x - 28 = 0.96\,y - 28.8$$

$$x = 0.96\,y - 28.8 + 28$$

$$\therefore \ x = 0.96\,y - 0.8$$

2. Regression Equation y on x

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

$$y - 30 = 0.83\,(x - 28)$$

$$y - 30 = 0.83\,x - 23.24$$

$$y = 0.83\,x - 23.24 + 30$$

$$\therefore \ y = 0.83\,x - 6.76$$

Q22. 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.

Jan-21, Q10 (OU)

OR

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

Solution :

May/June-19, Q9(a) (OU)

Calculation of Regression Equations

X	Y	dx = x − $\bar{x}$ ($\bar{x}$ = 10)	dx²	dy = y − $\bar{y}$ ($\bar{y}$ = 24)	dy²	dxdy
2	18	−8	64	−6	36	48
4	16	−6	36	−8	64	48
6	20	−4	16	−4	16	16
8	24	−2	4	0	0	0
10	22	0	0	−2	4	0
12	26	2	4	2	4	4
14	28	4	16	4	16	16
16	32	6	36	8	64	48
18	30	8	64	6	36	48
ΣX= 90	ΣY= 216	Σdx = 0	Σdx²= 240	Σdy= 0	Σdy²= 240	Σdxdy = 228

$$\therefore \bar{X} = \frac{\Sigma X}{N} = \frac{90}{9} = 10$$

$$\therefore \bar{Y} = \frac{\Sigma Y}{N} = \frac{216}{9} = 24$$

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

$$b_{yx} = \frac{\Sigma dxdy}{\Sigma dx^2} = \frac{228}{240} = 0.95$$

1. Regression Equation X on Y

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

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

$$X - 10 = 0.95Y - 22.8$$

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

$$\boxed{\therefore X = 0.95y + 12.8}$$

2. Regression Equation Y on X

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

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

$$Y - 24 = 0.95\,X - 9.5$$

$$Y = 0.95X - 9.5 + 24$$

$$\boxed{\therefore Y = 0.95x + 14.5}$$

3. Correlation Coefficient

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

$$= \pm\sqrt{0.95 \times 0.95}$$

$$= \pm\sqrt{0.9025}$$

$$= \pm 0.95$$

Since, both regression coefficients are positive, 'r' must be positive.

Hence, correlation coefficient = + 0.95

4. Regression Equation Y on X (When x = 6.2)

Substituting the value of Y when X= 6.2 Y will be,

$$Y = 0.95x + 14.5$$

$$Y = 0.95\,(6.2) + 14.5$$

$$Y = 5.89 + 14.5$$

$$\boxed{Y = 20.39}$$

PROBLEM ON ASSUMED MEAN REGRESSION METHOD

Q23. You are given the following information:

X	9	11	12	16	17	19	20	23
Y	4	5	6	7	8	10	12	15

Obtain two Regression Equations and find x when y = 25.

Solution :

July/Aug.-21, Q3 (KU)

Calculation of Regression Equations

X	Y	dx = x − $\bar{x}$ ($\bar{x}$ = 16)	dx²	dy = y − $\bar{y}$ ($\bar{y}$ = 8)	dy²	dxdy
9	4	-7	49	-4	16	28
11	5	-5	25	-3	9	15
12	6	-4	16	-2	4	8
16	7	0	0	-1	1	0
17	8	1	1	0	0	0
19	10	3	9	2	4	6
20	12	4	16	4	16	16
23	15	7	49	7	49	49
ΣX = 127	ΣY = 67	Σdx = −1	Σdx² = 165	Σdy = 3	Σdy² = 99	Σdxdy = 122

$\overline{X} = \frac{\Sigma x}{N} = \frac{127}{8} = 15.875$

$\overline{Y} = \frac{\Sigma y}{N} = \frac{67}{8} = 8.375$

Since mean values $\overline{X}$ and $\overline{Y}$ are non-integer values (decimal) deviations ae taken from assumed mean as 16 and 8 respectively.

1. **Regression Coefficient of x on y**

$$b_{xy} = \frac{n\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{n\Sigma dy^2 - (\Sigma dy)^2}$$

$$= \frac{8\times122 - (-1)(3)}{8\times99 - (3)^2}$$

$$= \frac{976+3}{792-9}$$

$$= \frac{979}{783}$$

$$= 1.2503$$

2. **Regression Equation X on Y**

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

X−15.875 = 1.2503 − (Y − 8.375)

X − 15.875 = 1.2503 Y − 10.4713

X = 1.2503 Y − 10.4713 + 15.875

∴X = 1.2503y + 5.4037

Now when, Y = 25 in that case X is

X = 1.2503 (25) + 5.4037

= 31.2575 + 5.4037

= 36.6612

3. **Regression coefficient of Y on X**

$$b_{yx} = \frac{n\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{n\Sigma dx^2 - (\Sigma dx)^2}$$

$$= \frac{8\times122-(-1)(3)}{8\times165-(-1)^2}$$

$$= \frac{976+3}{1,320-1}$$

$$= \frac{979}{1,319}$$

$$= 0.7422$$

4. **Regression Equation Y on X**

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

$Y-8.375=0.7422\,(X-15.875)$

$Y-8.375=0.7422\,X-11.7824$

$Y=0.7422\,X-11.7824+8.375$

$\therefore\quad Y=0.7422\,X-3.4074.$

Q24. Find the regression equations from the following data.

Age of Husbands	18	19	20	21	22	23	24	25	26	27
Age of Wives	17	17	18	18	19	19	19	20	21	22

Also calculate the correlation coefficient between the ages of the husbands and wives.

Solution :

July/Aug.-21, Q1 (MGU)

Calculation of Regression Equations

x	y	$dx = x-\overline{x}$ $(\overline{x}=22)$	dx^2	$dy = y-\overline{y}$ $(\overline{y}=19)$	dy^2	dxdy
18	17	-4	16	-2	4	8
19	17	-3	9	-2	4	6
20	18	-2	4	-1	1	2
21	18	-1	1	-1	1	1
22	19	0	0	0	0	0
23	19	1	1	0	0	0
24	19	2	4	0	0	0
25	20	3	9	1	1	3
26	21	4	16	2	4	8
27	22	5	25	3	9	15
$\Sigma x = 225$	$\Sigma y = 190$	$\Sigma dx = 5$	$\Sigma dx^2 = 85$	$\Sigma dy = 0$	$\Sigma dy^2 = 24$	$\Sigma dxdy = 43$

$\therefore\ \overline{X}=\frac{\Sigma x}{\Sigma n}=\frac{225}{10}=22.5$

$\therefore\ \overline{Y}=\frac{\Sigma y}{\Sigma y^2}=\frac{190}{10}=19$

$b_{xy}=\frac{\Sigma xy}{\Sigma y^2}=\frac{43}{24}=1.79$

$b_{yx}=\frac{\Sigma xy}{\Sigma y^2}=\frac{43}{85}=0.50$

1. **Regression Equation X on Y**

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

$x-22.5=1.79\,(y-19)$

$x=1.79y-34.01+22.5$

$\boxed{x=1.79\,y-11.51}$

2. **Regression Equation Y on X**

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

$y-19=0.50\ (x-22.5)$

$y-19=0.50\ x-11.25$

$y=0.50\ x-11.25+19$

$\boxed{y=0.50\ x+7.75}$

3. **Correlation Coefficient**

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

$r=0.50\times 1.79$

$r=0.895$

$\therefore\quad r=0.9460$

Q25. Given the following data, calculate the expected value of Y When X = 12. Find the Regression Equations.

	X	Y
Average	**7.6**	**14.8**
Standard Deviation	**3.6**	**2.5**

Where, r = 0.99.

Solution : **Sept./Oct.-21, Q10 (OU)**

Given, Mean $(\overline{x})=7.6$, Mean $(\overline{y})=14.8$

Standard Deviation (σ),

$x=3.6$, σ of $y=2.5$, $r=0.99$

Calculating the expected value of y when $x=12$ is,

$y-\overline{y}=r\dfrac{\sigma_y}{\sigma_x}(x-\overline{x})$

$y-14.8=0.99\times\dfrac{2.5}{3.6}(x-7.6)$

$y-14.8=0.69\ ((12)-7.6)$

$y-14.8=0.69\ (12)-5.24$

$y=0.69\ (12)-5.24+14.8$

$y=0.69\ (12)+9.56$

$y=0.69\ (12)+9.56$

$y=8.28+9.56=17.84$

$\therefore$ Expected value of y is 17.84

Q26. In a correlation study the following values are obtained

	X	Y
Mean	65	67
Standard Deviation	2.5	3.5
Coefficient of Correlation	0.8	

Find the two regression equations that are associated with the above values.

Solution : **July/Aug.-21, Q2 (MGU)**

Given that,

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

1. **Two Regression Equations**

(i) **Regression Equation of X on Y.**

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

$\Rightarrow x-65=0.8\left[\dfrac{2.5}{3.5}\right](y-67)$

$\Rightarrow x-65=0.5714\ (y-67)$

$\Rightarrow x-65=0.5714\ y-38.28$

$\Rightarrow x=0.5714\ y-38.28+65$

$\boxed{\Rightarrow x=0.5714y+26.72}$

(ii) **Regression Equation of Y on X.**

$y-\overline{y}=r\left[\dfrac{\sigma_x}{\sigma_x}\right](x-\overline{x})$

$\Rightarrow y-67=0.8\left[\dfrac{3.5}{2.5}\right](x-65)$

$\Rightarrow y-67=1.12\ (x-65)$

$\Rightarrow y-67=1.12\ x-72.8$

$\Rightarrow y=1.12\ x-72.8+67$

$\boxed{\Rightarrow y=1.12x+5.8}$

Q27. Calculate two regression equations from the following data.

X	4	6	8	13	13	12	14	15
Y	6	8	12	10	14	12	15	20

Solution : May/June-19, Q6(a) (MGU)

Calculation of Regression Equations

X	Y	dx (X = 11)	dx²	dy (Y = 12)	dy²	dxdy
4	6	−7	49	−6	36	42
6	8	−5	25	−4	16	20
8	12	−3	9	0	0	0
13	10	2	4	−2	4	−4
13	14	2	4	2	4	4
12	12	1	1	0	0	0
14	15	3	9	3	9	9
15	20	4	16	8	64	32
Σx = 85	Σy = 97	Σdx = −3	Σdx² = 117	Σdy = 1	Σdy² = 133	Σdxdy = 103

$$\overline{X} = \frac{\Sigma x}{N} = \frac{85}{8} = 10.625$$

$$\overline{Y} = \frac{\Sigma y}{N} = \frac{97}{8} = 12.125$$

Since mean values $\overline{X}$ and $\overline{Y}$ are non-integer values, deviations are taken from assumed mean as 11 and 12 respectively.

1. Regression Coefficient of Y on X

$$b_{yx} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dx^2 - (\Sigma dx)^2} = \frac{8 \times 103 - (-3)(1)}{8\times117 - (-3)^2}$$

$$= \frac{824+3}{936-9} = \frac{827}{927}$$

$$= 0.8921$$

2. Regression Equation of Y on X

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

$$Y - 12.125 = 0.8921(X - 10.625)$$

$$Y - 12.125 = 0.8921X - 9.4786$$

$$Y = 0.8921X - 9.4786 + 12.125$$

$$Y = 0.8921X + 2.6464$$

3. Regression Coefficient of X on Y

$$b_{xy} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{N\Sigma dy^2 - (\Sigma dy)^2}$$

$$= \frac{8\times103 - (-3)(1)}{8\times133 - (1)^2} = \frac{824+3}{1064-1} = \frac{827}{1063}$$

$$= 0.778$$

4. Regression Equation of X on Y

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

$$X - 10.625 = 0.778\,(Y - 12.125)$$

$$X - 10.625 = 0.778Y - 9.4332$$

$$X = 0.778\,Y - 9.4332 + 10.625$$

$$X = 0.778Y + 1.1918$$

PROBLEM ON ACTUAL OBSERVATION REGRESSION METHOD OR VALUE BASED METHOD

Q28. 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.

Solution : May/June-18, Q9(b) (OU)

Given ,

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

(i) Two Regression Equations

Regression Equation of X on Y

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

$$\bar{x} = \frac{\Sigma x}{N} = \frac{56}{8} = 7$$

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

$$= \frac{8 \times 364 - (56)(40)}{8 \times 256 - (40)^2}$$

$$= \frac{2,912 - 2,240}{2,048 - 1,600}$$

$$= \frac{672}{448}$$

$$= 1.5$$

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

$$x - 7 = 1.5(y - 5)$$

$$x - 7 = 1.5y - 7.5$$

$$x = 1.5y - 7.5 + 7$$

$$x = 1.5y - 0.5$$

Regression Equation of Y on X

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

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

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

$$= \frac{8 \times 364 - (56)(40)}{8 \times 524 - (56)^2}$$

$$= \frac{2,912 - 2,240}{4,192 - 3,136}$$

$$= \frac{672}{1056}$$

$$= 0.6364$$

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

$$y - 5 = 0.6364(x - 7)$$

$$y - 5 = 0.6364x - 4.4548$$

$$y = 0.6364x - 4.4548 + 5$$

$$y = 0.6364x + 0.5452$$

(ii) Correlation Coefficient

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

$$= \pm\sqrt{0.6364 \times 1.5}$$

$$= \pm\sqrt{0.9546}$$

$$= \pm 0.98$$

PROBLEM ON GROUPED DATA REGRESSION METHOD OR STEP DEVIATION METHOD

Q29. Following is the distribution of students according to their heights and weights.

Heights (in inches)	Weight in (Lbs)			
	90-100	100-110	110-120	120-130
50-55	4	7	5	2
55-60	6	10	7	4
60-65	6	12	10	7
65-70	3	8	6	3

Calculate,

(i) The two coefficients of regression

(ii) Obtain the two regression equations.

Solution :

Let height be 'x' and weight be 'y'

(i) The Two Coefficients of Regression

(a) Regression coefficient of x on y

$$b_{xy} = \frac{N.\Sigma fdxdy - \Sigma fdx\ \Sigma fdy}{N.\Sigma fd^2 y - (\Sigma fdy)^2} \times \frac{i_x}{i_y}$$

(b) Regression coefficient of y on x

$$b_{yx} = \frac{N.\Sigma fdxdy - \Sigma fdx\ \Sigma fdy}{N.\Sigma fd^2 x - (\Sigma fdx)^2} \times \frac{i_y}{i_x}$$

Where,

i_y = Difference between class intervals of weight
$= 100 - 90 = 10$

i_x = Difference between class intervals of height
$= 55 - 50 = 5$

Calculations of Coefficient Regression

Height (x)	Mid - value	M.V. dx \ dy	Weight (y) 90-100 95 −1	100-110 105 0	110-120 115 1	120-130 125 2	f	fdx	f₂dx	fdxdy
50-55	52.5	−1	4 4	7 0	5 −5	2 −4	18	−18	18	−5
55-60	57.5	0	6 0	10 0	7 0	4 0	27	0	0	0
60-65	62.5	1	6 −6	12 0	10 10	7 14	35	35	35	18
65-70	67.5	2	3 −6	8 0	6 12	3 12	20	40	80	18
		f	19	37	28	16	N=100	57	133	31
		fdy	−19	0	28	32	41			
		fdy²	19	0	28	64	111			
		fdxdy	−8	0	17	22	31			

Here, $\sum fdxdy = 31, \sum fdx = 57, \sum fdy = 41, N = 100$

$\sum fdx^2 = 133$ and $\sum fdy^2 = 111$

(a) Regression Coefficient of x on y

$$b_{xy} = \frac{N.\Sigma fdxdy - \Sigma fdx.\Sigma fdy}{N.\Sigma fdy^2 - (\Sigma fdy)^2} \times \frac{i_x}{i_y} = \frac{100\times 31 - (57)(41)}{100\times 111 - (41)^2} \times \frac{5}{10}$$

$$= \frac{3100 - 2337}{11100 - 1681} \times \frac{1}{2} = \frac{763}{9419} \times \frac{1}{2}$$

$$= 0.0810 \times 0.5 = 0.0405 \cong 0.041$$

(b) Regression Coefficient y on x

$$b_{yx} = \frac{N.\Sigma fdxdy - \Sigma fdx.\Sigma fdy}{N.\Sigma fdx^2 (\Sigma fdx)^2} \times \frac{i_y}{i_x}$$

$$= \frac{100 \times 31 - (57)(41)}{100 \times 133 - (57)^2} \times \frac{10}{5}$$

$$= \frac{763}{13300 - 3249} \times \frac{10}{5} = \frac{763}{10051} \times 2$$

$$= 0.0759 \times 2 = 0.1518 \simeq 0.152$$

(ii) Two Regression Equations

(a) Regression Equation of x on y

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

$$x - 60.35 = 0.041(y - 109.1)$$

$$x - 60.35 = 0.041y - 4.473$$

$$x = 0.041y - 4.473 + 60.35$$

$$x = 0.041y + 55.88$$

(b) Regression Equation of y on x

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

$$y - 109.1 = 0.152(x - 60.35)$$

$$y = 0.152\,x - 9.17 + 109.1$$

$$y = 0.152\,x + 99.93$$

Working Notes

Calculation of $\bar{x}$ and $\bar{y}$

$$\bar{x} = A + \frac{\Sigma fdx}{N} \times i_x$$

Where,

A= Assumption

$$\bar{x} = 57.5 + \frac{57}{100} \times 5$$

$$= 57.5 + 2.85$$

$$\bar{x} = 60.35$$

$$\bar{y} = A + \frac{\Sigma fdy}{N} \times i_y$$

$$\bar{y} = 105 + \frac{41}{100} \times 10$$

$$\bar{y} = 105 + 4.1 = 109.1$$

EXERCISE AND PRACTICE QUESTIONS

SHORT QUESTIONS

THEORY

Q1. Define Regression Analysis. [Refer, Q1]

Q2. Features of Regression Coefficients. [Refer, Q2] May/June-19, Q1(MGU)

Q3. Write about principle of least squares and standard error of estimate. [Refer, Q4]

Q4. Write three limitations of regression analysis. [Refer, Q5] May/June-18, Q1(a) (KU)

PROBLEMS

Q5. If $r = 0.6$; $\sigma_x = 5$, $\sigma_y = 7$, find b_{xy} and b_{yx}. [Refer Similar, Q8]

(Ans: $b_{xy} = 0.4284$, $b_{yx} = 0.84$)

Q6. If $\gamma = 0.8$, $\sigma_x = 3$ and $\sigma_y = 4$, find the b_{xy} and b_{yx}. [Refer Similar, Q9]

(Ans: $b_{xy} = 0.6$, $b_{yx} = 1.067$)

Q7. Co-efficient of correlation = 0.60, $\sigma_x = 3$, $\sigma_y = 4$, $x = 10$, $y = 20$ find regression equation y on x. [Refer Similar, Q10]

(Ans: $b_{yx} = 0.79$ or 0.8 ; Y on X; $Y = 0.8X + 12$)

Q8. 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. [Refer Similar, Q28]

(Ans: +0.70)

Q9. If regression equation of X on Y = 0.268 and of Y on X = 0.5, find coefficient of correlation. [Refer Similar, Q27]

(Ans: +0.259)

Q10. If the correlation coefficient (r) = 0.86 and regression coefficient of X on Y = 1.2, find the regression coefficient of Y on X. [Refer Similar, Q28]

(Ans: 0.6163)

ESSAY QUESTIONS

THEORY

Q1. Define and explain regression analysis? Explain regression variables and types of regression. [Refer, Q11]

Q2. Write the relation between correlation and regression. [Refer, Q13] May/June-18, Q2(a) (KU)

Q3. What do you mean by linear and nonlinear regression? Distinguish between them. [Refer, Q14]

Q4. Define regression and what are the differences between correlation and regression. [Refer, Q15]

May/June-18, Q9(a) (OU)

PROBLEMS

Q5. From the following data obtain the regression equation of x on y and also that of y on x. [Refer Similar, Q22]

x	6	2	10	4	8
y	9	11	5	8	7

(Ans: x = 1.3y + 16.4; y = – 0.65 x + 11.9)

Q6. From the following data obtain the two regression equations and calculate the correlation co-efficient. [Refer Similar, Q22]

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Estimate the value of y which should correspond on an average to X = 6.2

(Ans: y = 13.14)

Q7. From the data given below find out,

(a) Co-efficient of correlation between the ages of husbands and the ages of wives.

(b) The two regression equations.

(c) The expected age of husband when wife's age is 14.

(d) The expected age of wife when husband's age is 35. [Refer Similar, Q22, Q23]

Age of Husband (in years)	22	23	23	24	26	27	27	28	30	30
Age of Wife (in years)	18	20	21	20	21	22	23	24	25	26

(Ans: x = 1.1y + 1.8, y = 0.8 x + 1.2, r = 0.94, x = 17.2, y = 29.2)

Q8. The following data about the sales and advertisement expenditure of a firm given below,

	Sales (₹ Crore)	Adv.Exp (₹ Crore)
Mean	40	6
S.D	10	1.5
Coefficient correlation		r = + 0.9

(i) What should be the advertisement expenditure if the firm proposed a sales targets of 50 crores of rupees?

(ii) Estimate the likely sales for a proposed advertisement expenditure of 12 crores of rupees. [Refer Similar, Q25]

(Ans: x = 0.135y + 0.6; y = 6x + 4; y = 304; x = 2.222)

Q9. Following data are given for marks in english (x) and marks in maths (y) at a certain examination.

	x	y
Mean marks	39.5	47.5
S.D of marks	10.8	16.8
Co-efficient correlation is	0.42	

Find the regression equation of y on x and estimate the marks in maths when marks in english is 50. [Refer Similar, Q25]

(Ans: $y = 0.653\,x - 21.70$; $y = 10.95$)

Q10. 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. [Refer Similar, Q27] Oct./Nov.-16, Q12(b) (OU)

(Ans: $x = 0.45y + 1$; $y = 0.799x + 12.01$)

Q11. Following are the marks in Statistics and English in an Annual Examination.

	Statistics (X)	English (Y)
Mean	40	50
Standard Derivation	10	16
Coefficient 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.[Refer Similar, Q25]

(Ans: $X = 0.3125y + 24.375$, $Y = 0.8x + 18$; $X = 34$; $Y = 58$.) May/June-19, Q9(b) (OU)

Q12. Calculate two regression equations from the following data. [Refer Similar, Q22]

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

(Ans: $X = 0.6y + 1.6$, $Y = 0.6x + 1.6$) May/June-19, Q6(b) (MGU)

Q13. Find out two regression equations from the following data:[Refer Similar, Q22]

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

(Ans: $X = 0.9y - 0.6$, $Y = 0.9x + 1.3$) May/June-18, Q2(b) (KU)

INTERNAL ASSESSMENT/EXAM

1 Multiple Choice

1. _________ is the study of mathematically measuring the average relationships, if it exists between two or more variables. [ ]
   (a) Correlation analysis (b) Regression analysis
   (c) Correlation coefficient (d) Correlation ratios
2. The variable which influences the values of other variable, [ ]
   (a) Dependent variable (b) Regression
   (c) Independent variable (d) Regression coefficient
3. The variable whose value is influenced or is to be predicted, [ ]
   (a) Dependent variable (b) Regression
   (c) Regression coefficient (d) Independent variable
4. If the regression curve is a straight line, then the regression is termed as _________ regression. [ ]
   (a) Simple (b) Multiple
   (c) Non-linear (d) Linear
5. Standard error of estimate is also known as _________. [ ]
   (a) Standard error (b) Standard error of regression
   (c) Standard error of prediction (d) None of the above
6. If the curve of the regression is not a straight line, then the regression is termed as _________ regression. [ ]
   (a) Simple (b) Multiple
   (c) Non-linear (d) Linear
7. _________ uses the principle of least squares to give the best fit line for estimating the value of one variable given the value of another variable. [ ]
   (a) Standard error of estimate (b) Regression coefficient
   (c) Regression equation (d) Lines of regression
8. _________ consists of minimizing the sum of the squares of the residuals or error of estimates. [ ]
   (a) Principle of least square (b) Lines of regression
   (c) Regression equation (d) Regression coefficient
9. If the regression lines are expressed in an algebra terms, then it is called as _________. [ ]
   (a) Regression (b) Regression equation
   (c) Regression coefficient (d) None of the above
10. For two variables X and Y, there are _________ lines of regression. [ ]
   (a) One (b) Three
   (c) Two (d) Four

II Fill in the Blanks

1. ________ studies nature of relationship between the variables.
2. Regression takes its name from studies made by ________.
3. ________ are independent of the change of the origin but not the change of scale.
4. Regression analysis involves two types of variables i.e., ________ and ________ variables.
5. The least square regression line always passes through ________.
6. Regression analysis is also used in ________.
7. The equation of ________ is commonly use to predict the value of Y for a given value of X.
8. ________ indicate the average values of one variable for a given or known values of other variable.
9. When the degree of correlation is high, then the ________ of regression will be close to each other.
10. There are two ways of forming regression equations, i.e., ________ and regression coefficient.

KEY

I. Multiple Choice

1. (b)
2. (c)
3. (a)
4. (d)
5. (b)
6. (c)
7. (d)
8. (a)
9. (b)
10. (c)

II. Fill in the Blanks

1. Regression
2. Sir Francis Galton
3. Regression coefficient
4. Dependent and independent
5. $(\bar{x}, \bar{y})$
6. Optimization
7. Regression line
8. Straight line
9. Two lines
10. Normal equation.

III Very Short Questions and Answers

Q1. Define Regression Analysis.

Answer :

According to Ya-Lun 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 do you mean by Linear Regression?

Answer :

Linear Regression is a form of regression which is used for modeling the relationship between scalar variables like '*X*' and '*Y*'. Under linear regression, linear functions are used to model the data and the unknown parameters of models are estimated from the data. Therefore, these models are known as Linear Models.

Q3. Write a short note on lines of regression.

Answer :

"The regression lines show the average relationship between two variables". *–Galton*

"The device used for estimating the value of one variable from the value of the other consists of line through the points, drawn in such a manner as to represent the average relationship between the two variables. Such a line is called the line of regression".

–J.R. Stockton

Q4. What do you mean by non-linear regression?

Answer :

Non-Linear Regression refers to the explained variable (dependent variable) which 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. Write about principle of least square.

Answer :

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.

UNIT 2

INDEX NUMBERS

SYLLABUS

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.

LEARNING OBJECTIVES

- ✓ Introduction, Characteristics, Uses, Limitations and Types of Index Numbers.
- ✓ Problems in the Construction of Index Numbers.
- ✓ Methods of Constructing Index Numbers (Laspeyre's, Paasche, Marshall-Edgeworth and Fishers).
- ✓ Tests of Consistency of Index Numbers.
- ✓ Base Shifting, Splicing and Deflating of Index Numbers.

INTRODUCTION

Index numbers are numerical values or devices or series of numbers that measure the average changes in price or quantity over a period of time. Index numbers are expressed in percentages. They are calculated by dividing current value by the base value and then multiplied by 100. Index numbers do not have units.

Index numbers are helpful in measuring fluctuations in the economic conditions. They act as a barometer and measure the pressure of economic and business behaviour.

The various methods of constructing index numbers are: Weighted Price Indexes or Indices, Simple Price Indexes or Indices (Unweighted Price Indexes or Indices).

Laspeyre's index method was introduced by a famous statistician named "Laspeyre". In this method, prices of all items or commodities are weighed by the quantity, consumed both in the base and the current year.

In Paasche's index method, prices are weighted by the quantity consumed in the current year, rather than considering quantity consumed in base year.

In Marshall-Edgeworth's index method, both the current year as well as base year quantities are considered to calculate the index.

A statistician named "Fisher" introduced this method, which is a geometric mean of Laspeyre's and Paasche's methods. Fisher's ideal index takes into consideration both the base year and current year quantities.

The various tests of consistency of index numbers are as follows, Unit Test, Time Reversal Test, Factor Reversal Test, Circular Test.

Deflating refers to the process of making allowances for the impact of changing prices. An increase in price level leads to a decrease in purchasing power of money.

PART-A
SHORT QUESTIONS AND ANSWERS

Q1. Define Index Numbers. Jan-21, Q2(OU)

Answer :

According to Maslow, "Index number is a numerical value characterizing the change in complex economic phenomena over a period of time or space".

According to Croxton and Cowden, "Index numbers are devices measuring differences in the magnitude of a group of related variables".

According to Horace Secrist, "Index numbers are series of numbers by which changes in the magnitude of a phenomenon are measured from time to time or place to place".

Based on the above definitions, Index numbers can be defined as, a specialized average designed to measure the change of related variable over a period of time".

Q2. Importance of Index Numbers May/June-18, Q2 (OU)

OR

Advantages of Index Numbers (Sept/Oct-21, Q2 (OU) | March/April-17, Q8 (OU))

OR

Explain the uses of Index Numbers. March/April-14, Q7 (OU)

Answer :

Following are the importance, advantages or uses of index numbers,

1. Helps in Measuring Economic Conditions

Index numbers are helpful in measuring fluctuations in the economic conditions. They act as a barometer and measure the pressure of economic and business behaviour. For example, the composite index numbers of foreign exchange reserves, industrial output and bank deposits can act as an economic barometer.

2. Policy Formulation

Price index reveals the fluctuations in prices over a period of time. Movement in prices directly effects the business as well as economic operations. Price index guides the business organizations and governmental bodies in formulating a new policy or modifying the existing policy in order to meet the issues that may arise due to price fluctuations.

3. Deflating Various Values

Index numbers can also be used for deflating i.e., they are useful to adjust original data for price changes. For example, various values such as national income of the population.

4. Inflationary and Deflationary Tendencies

Now-a-days index numbers are also being widely used for measuring the inflation and deflation. They are acting as indicators for inflationary and deflationary tendencies.

Q3. Types of Index Numbers

Answer : (May/June-19, Q2 (OU) | Oct./Nov.-14, Q4 (OU))

Index numbers are classified into three types as follows,

1. Price Index Numbers

This method is used to compare the prices of various commodities of one year with that of the another year. It is further classified into two types,

(i) Wholesale Price Index Numbers

(ii) Retail Price Index Numbers.

2. Quantity Index Numbers

Quantity index numbers measure the variation in the volume of goods manufactured or consumed or distributed in quantitative terms. It measures the change in quantitative terms like, increase or decrease in volume of production, construction or employment.

3. Value Index Numbers

Value index numbers are used to measure the changes that takes place in the total value of the variable. The total value of the variable is obtained by multiplying price (p) with quantity (q).

Value = Price × Quantity

Q4. Marshall EdgeWorth Method

Answer :

May/June-19, Q2 (MGU)

Marshall-Edgeworth's index method refers to the calculation of index for both the current year as well as base year quantities. The formula for Marshall-Edgeworth's Index method is as follows,

$$I_p\,(ME) = \frac{\Sigma(q_0+q_1)P_1}{\Sigma(q_0+q_1)P_0}\times 100$$

$$P_{01} = \left(\frac{\Sigma P_1 q_0 + \Sigma P_1 q_1}{\Sigma P_0 q_0 + \Sigma P_0 q_1}\right)\times 100$$

The following are the advantages of Marshall-Edgeworth's index method,

1. It is easy to understand and calculate.
2. It uses current and base year's quantities and prices.
3. It satisfies both the unit and time reversal tests of consistency.

Q5. What is "Cost of living Index"?

Answer :

May/June-18, Q1(b) (KU)

Consumer price index numbers is also known as cost of living index number. It is used to measure the purchasing power of a particular class of people in relation to the changes in retail prices. In other words, it studies how price variations effect the cost of living or purchasing power of a group of people.

While constructing cost of living index number, a particular section of society is selected like [Rich, middle, poor] and a study is conducted to know how price variations effect the consumption levels of that section. Based on such information Cost of Living Index Number (CLIN) is constructed.

Q6. 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 2016 (₹)	40	60	85	25	30
Price 2017 (₹)	60	90	125	30	40

Answer :

Jan.-21, Q3 (OU)

Computation of Simple Aggregative Method

Commodity	Price in (₹)	
	2016 (P_0)	2017 (P_1)
P	40	60
Q	60	90
R	85	125
S	25	35
T	30	40
Total	$\Sigma P_0 = 240$	$\Sigma P_1 = 350$

Calculation of index number by using simple aggregate method,

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100$$

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

$$= 1.4583 \times 100$$

$$\therefore P_{01} = 145.83.$$

Q7. Following are the prices of commodities is 1970 and 1975. Calculate a Price Index based on Price Relatives using the Arithmetic Mean.

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

Answer : (Sept./Oct.-21, Q8 (OU) | May/June-18, Q3 (OU))

Calculation of Price Index Based on Price Relatives Method by Using Arithmetic Mean

Commodities	Prices (₹)		Price Relatives $\left[\frac{P_1}{P_0}\right]100$
	2015 (P_0)	2016 (P_1)	
A	45	55	122.22
B	60	70	116.67
C	20	30	150
D	50	75	150
E	85	90	105.88
F	120	130	108.33
N = 6			$\Sigma\left[\frac{P_1}{P_0}\right]100 = 753.1$

$$P_{01} = \frac{\Sigma\left[\frac{P_1}{P_0}\right]100}{N}$$

$$= \frac{753.1}{6}$$

$$= 125.517$$

Q8. 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

Answer : May/June-19, Q3 (OU)

Calculation of Index Number by Average Price Relative Method Using Arithmetic Mean

Commodity	Prices		Price Relatives
	2017 (p_0)	2018 (p_1)	$\left[\frac{p_1}{p_0}\right]100$
P	2	4	200
Q	6	8	133.3
R	10	15	150
S	5	5	100
T	12	8	66.7
N = 5			$\Sigma\left[\frac{p_1}{p_0}\right]100 = 650$

$$P_{01} = \frac{\Sigma\left[\frac{P_1}{P_0}\right]100}{N} = \frac{650}{5} = 130$$

Q9. Compute price index by weighted average of price relatives method using arithmetic mean.

Commodities	A	B	C	D	E
Price in base year	10	6	14	22	18
Quantity in base year	160	180	120	40	80
Price in current year	16	8	14	28	24

Answer : March/April-15, Q8 (OU)

Calculation of Price Index by Weighted Average of Price Relatives Method using Arithmetic Mean

Commodities	p_0	q_0	p_1	p_0q_0 V	$\frac{p_1}{p_0}\times 100$ P	PV
A	10	160	16	1,600	160	2,56,000
B	6	180	8	1,080	133.33	1,43,996.4
C	14	120	14	1,680	100	1,68,000
D	22	40	28	880	127.27	1,11,997.6
E	18	80	24	1,440	133.33	1,91,995.2
				ΣV = 6,680		PV = 8,71,989.2

$$p_{01} = \frac{\Sigma PV}{\Sigma V} = \frac{8{,}71{,}989.2}{6680}$$

$= 130.537$.

∴ It means that, there is a increase in prices over base level by 130.537%.

PART-B

ESSAY QUESTIONS AND ANSWERS

2.1 INDEX NUMBERS - INTRODUCTION, CHARACTERISTICS, USES, LIMITATIONS AND TYPES

Q10. Define Index Number. What are its features and uses? May/June-19, Q10(a) (OU)

OR

What is Index Number? Write importance of index numbers.

(Refer Only Topics: Index Numbers, Uses/Importance of Index Numbers)

May/June-18, Q3(a) (KU)

Answer :

Index Numbers

According to Maslow, "Index number is a numerical value characterizing the change in complex economic phenomena over a period of time or space".

According to Croxton and Cowden, "Index numbers are devices measuring differences in the magnitude of a group of related variables".

According to Horace Secris, "Index numbers are series of numbers by which changes in the magnitude of a phenomenon are measured from time to time or place to place".

In simple terms, Index numbers are numerical values or devices or series of numbers that measure the average changes in price or quantity over a period of time. Index numbers are expressed in percentages and also they do not have units.

Features/Characteristics of Index Numbers

The features or characteristics of index numbers are as follows,

1. Measure the Change in Percentages

Index numbers measure the change in price or quantity in terms of percentages such as, 10%, 20%, 15%, 25% and so on. Increase or decrease in value is represented by one single figure. Like 10% increase in sales from that previous year to current year, 30% decrease in profits when compared to that of the last year and so on.

2. Specialized Averages

A single figure known as "Average" is used for representing the characteristics of the complete set of data. Average acts as a basis for comparing different data sets with each other, when they have common unit of measurement of observations. When data sets do not have common unit of measurement, specialized averages of index numbers are used for the comparison.

3. The Measured Changes cannot be Observed Directly

Index numbers do not measure the changes directly but it studies the relative changes or variations in factors resulting to changes-like measuring changes in Export-Imports related to factors such as available raw materials, technology, competitors etc.

4. Measures Changes in Relation to Time or Place

Index numbers measure change by comparing the values at different time periods or at different places like standard of living at one place is being compared with standard of living of the other place. Sales or revenue of current year is compared with that of the previous years sales or revenue.

Uses/Importance of Index Numbers

For answer refer Unit-II, Page No. 35, Q.No. 11, Topic: Uses/Importance of Index Numbers.

Q11. What is the importance and limitations of index numbers? Explain.

Answer : May/June-19, Q7(b) (MGU)

Uses/Importance of Index Numbers

The uses/importance of index numbers are as follows,

1. Helps in Measuring Economic Conditions

Index numbers are helpful in measuring fluctuations in the economic conditions. They act as a barometer and measure the pressure of economic and business behaviour. For instance, the composite index numbers of foreign exchange reserves, industrial output and bank deposits can act as an economic barometer.

2. Measuring Purchasing Power

Purchasing power of a group or class can be easily measured with the help of the index numbers. As purchasing power is related with a group of people, price index is required for providing overall view of the purchasing power of the group. By considering earnings and expenses on purchases of a particular group. For example, index numbers can measure the fluctuations (increase or decrease) in purchasing power of that group, over a period of time.

3. Reveal Trend and Tendencies

Index numbers measure the average changes in phenomenon, by taking into account the current year and basic year values. The price index gives the fluctuations in specific years, through which trends of the phenomenon can be easily represented in graph. Even conclusions can also be framed by analyzing these trends.

4. Policy Formulation

Price index reveals the fluctuations in prices over a period of time. Movement in prices directly effects the business as well as economic operations. Price index guides the business organizations and governmental bodies in formulating a new policy or modifying the existing policy in order to meet the issues that may arise due to price fluctuations.

5. Deflating Various Values

Index numbers can also be used for deflating various values such as national income of the overall population.

6. Inflationary and Deflationary Tendencies

Now-a-days index numbers are also being widely used for measuring the inflation and deflation. They are acting as indicators for inflationary and deflationary tendencies.

Limitations of Index Numbers

The following are the various limitations of index numbers,

1. Difficulty in Selection of Data

It is difficult to include each and every item in the construction of index as such construction takes into consideration of only the selected sample data.

2. Random Sampling

In the construction of index numbers, random sampling is hardly used as selecting a sample from huge population as random sampling procedure is not practical. Inspite of selecting the samples carefully, some errors might take place in the construction of indices. Therefore, efforts must be made to reduce such errors.

3. Quality of Product

Under index numbers, the quality of products should remain same during a period of time. Dissimilarity in quality of the products means dissimilarity in prices of products which makes the comparison during a period of time less authentic in nature.

4. Methods of Construction

Index numbers might be constructed by using different methods which gives different results. Further, it requires a selection of an appropriate formula, otherwise, it would result in problems of comparison of results. Due care should be taken while selecting a formula and same formula should be used over a period of time.

5. Misuse

It might be misused by some dishonest capitalists who can show less profits in the company. Current year profits are compared with the record year profits to show less profits in the company. Likewise, fraudulent trade compares the current year prices with the year where prices were very high to show current year prices as very high prices.

6. Inaccurate Information

Non- availability of accurate and sufficient information is one of the limitation in the construction of index numbers.

7. Data collection

Data needs to be collected from geographically scattered locations.

8. Fake results

The chances of fake results are high. It does not provide appropriate results.

9. Changes in price

The price of one commodity might differ from the prices of another commodities due to different prices fixed by the wholesalers or retailers.

Q12. Explain the various types of Index Numbers.

Answer :

The various types of index numbers are as follows,

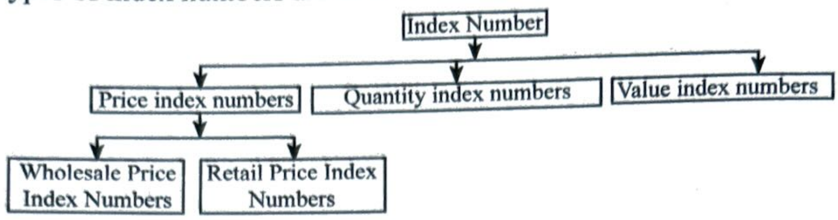


Figure: Types of Index Number

1. Price Index Numbers

Price index number method is used to compare the prices of various commodities of one year with that of the another year. It is further classified into two types. They are as follows,

(i) Wholesale Price Index Numbers

Wholesale price indices are used to study the changes that takes place in the general price level of a country.

(ii) Retail Price Index Numbers

Retail Price Index Number is used to study the changes that takes place in retail prices of various commodities like fruits, vegetables, rice, wheat etc.

2. Quantity Index Numbers

Quantity index numbers measure the variation in the volume of goods manufactured or consumed or distributed in quantitative terms. It measures the change in quantitative terms like, increase or decrease in volume of production, construction or employment.

3. Value Index Numbers

Value index number are used to measure the changes that takes place in the total value of the variable. The total value is obtained by multiplying price (p) with quantity (q).

$$\text{Value} = \text{Price} \times \text{Quantity}$$

2.2 PROBLEMS IN THE CONSTRUCTION OF INDEX NUMBERS

Q13. Explain the problems in the construction of Index Numbers.

Answer : July/Aug.-21, Q1(b) (KU)

While constructing index numbers the following problems must be taken into consideration,

1. Purpose/Object of Index

While constructing index number the purpose/object should be clearly decided. For example, if consumer's cost of living index number is used to measure standard of living of poor families then in such case utmost care should be taken that it should not include any other data of middle class and rich class in it. Otherwise, it would result in confusion and wastage of time and accurate results would not be obtained.

2. Selection of a Base Period

A base period is selected to compare the two periods. It might be either a year, a month or a day. Index for base period should be always taken as 100. While selecting a base period the following points should be taken into account,

(i) A base period selected should be free from abnormalities such as earthquakes, wars, famines, depression booms and so on. But some times selecting a base period which is free from all abnormalities becomes a difficult task.

(ii) A selected base year should not be too far from the current year.

(iii) While constructing price index number a decision has to be taken whether to proceed with fixed base or chain base index. Under fixed base method a fixed year is selected for the entire series. Whereas in chain method, price are compared with their preceding year instead of the fixed year. However, chain base method gives better results when compared to the fixed base method.

3. Number of Commodities to be Included

In construction of price index number only those commodities are taken into consideration which best represent the tastes, preferences of the consumers and habits of the people for whom index is constructed.

For example, if price index number is used to measure the monthly budget of a particular family. Then, in such a case it should include only items like, Clothing, fuel and light. Apart from these no other items should be included. The selection of number of commodities to be included is decided based upon the purpose of the index.

4. Price Quotations

After the selection of commodities the next problem which arises is collection of the accurate price quotations for such commodities. As prices of commodities differ from place to place and from shop to shop within the market it becomes impossible to get price quotations from all such places where commodity is manufactured. So respective persons and places should be selected. The data should be obtained from reliable sources like journal, magazine newspapers and government organizations. Therefore, in order to ensure uniformity in prices, two methods of quoting prices are followed. They are,

(i) Money prices and

(ii) Quantity prices.

Price quotations are available in two forms i.e., wholesale and retail. A decision regarding whether to select wholesale prices or retail prices are made based upon the objective, purpose of index number.

5. Choice of an Average (Two Contradictory Statements are Given)

Before constructing index numbers, selecting an average is mandatory. In practice, median, mode and mean are not used for constructing the price index. Among all averages, geometric mean is considered as the most suitable one for construction of index numbers because of the following reasons,

(i) In the construction of index numbers focus is laid on finding out the ratios of change.

(ii) Results obtained by using geometric mean are reversible. Thereby, it facilitates base shifting.

6. Selection of an Appropriate Formula

Selection of an appropriate formula is made on the basis of the purpose of index and available data. According to Professor, Irving Fisher, an index is considered as appropriate, when it is both time reversal test and factor reversal test. However, in practice, there is no particular formula which can be considered as appropriate in situations.

7. Selection of Appropriate Weights

It is important to select appropriate weights for the items to measure the relative importance of different items in the construction of index numbers. Indices can be broadly categorized into two types. They are as follows,

(i) Weighted indices and

(ii) Unweighted indices.

In case of weighted indices, particular weights are assigned to the items. But in case of unweighted indices no specific weights are assigned to the items.

Weights can be assigned to the items by two methods. They are,

(a) Implicit method and

(b) Explicit method.

Ultimately, the decision needs to be taken regarding the types of weights like whether it should select fixed or fluctuating weights. Selection of appropriate weights is very important and also a very difficult task.

2.3 METHODS OF CONSTRUCTING INDEX NUMBERS - SIMPLE AND WEIGHTED INDEX NUMBERS

Q14. Define Index Number. Explain the various methods of constructing index numbers?

July/Aug.-21,Q4 (KU)

Answer :

Index Number

According to Croxton and Cowden, "Index numbers are devices measuring differences in the magnitude of a group of related variables".

According to Horace Secris, "Index numbers are series of numbers by which changes in the magnitude of a phenomenon are measured from time to time or place to place".

Methods of Constructing Index Numbers

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

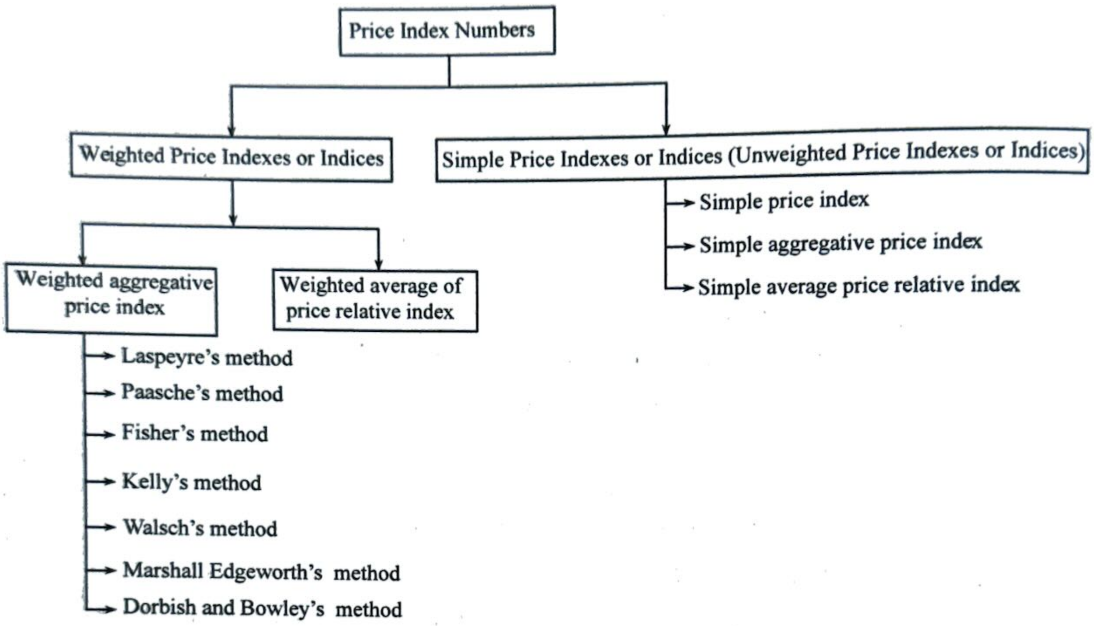


Figure: Types and Methods of Price Index Numbers

1. Weighted Price Indexes or Indices

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. They are,

(a) Weighted Aggregative Price Index

In a weighted aggregative 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) Marshall Edgeworth's method

(vii) Dorbish and Bowley's method.

(b) Weighted Average of Price Relative Index

In weighted average of price relative index, value of each commodity or item related with the calculation of composite index is ascertained by multiplying the price of each item with its quantity consumed. Quantity consumed is considered for computing the weighted average of price relative. The formula for weighted average of price relative index is as follows,

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

$$= \frac{\Sigma PV}{\Sigma V}$$

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

Where,

$V (= p_0 q_0)$ = Base prices and quantities determining values.

$P\left(= \frac{p_1}{p_0} \times 100\right)$ = Price relative

This formula is equivalent to the formula of Laspeyre's index formula,

If 'V' is taken as $p_0\, q_1$, then the formula would,

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

Then it would be equal to Paasche's index method.

2. Simple/Unweighted Price Indexes or Indices

The simple or unweighted indexes or indices include the following methods,

(a) Simple/Single Price Index

Single price index is computed by dividing the current year price of the commodity with its base year price. It is a percentage ratio which represents the comparison of a particular commodity price. The general formula used for single price index is as follows,

$$\text{Single price index in period 'n'} = \frac{p_n}{p_0} \times 100$$

Where,

p_n = Price of the commodity in the n^{th} year

P_0 = Price of the commodity in the base year.

(b) Simple Aggregate Price Index

In simple aggregate price index, the sum of current year prices of various commodities is divided with the sum of base year prices of that various commodities. The formula is given as follows,

Simple aggregate price index,

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

Where,

p_1 = Unit price of a current year prices of all commodities

p_0 = Sum of base year prices of all commodities.

(c) Simple Average Price Relative Index

This method is an improvement over the aggregate price method. The formula for this method is,

$$P_{01} = \Sigma\left(\frac{P_1}{P_0} \times 100\right)$$

Where,

n = Number of commodities included in the computation of the index.

Q15. Explain the various methods of weighted aggregate price index.

Answer :

The various methods used in weighted aggregate price index are as follows,

1. Laspeyre's Index Method

Laspeyre's index method was introduced by a famous statistician named "Laspeyre". In this method, prices of all items or commodities are weighted by the quantity, consumed both in the base and the current year. Index of the different periods can be directly compared with each other in this method, as the index number relies upon the same base price and quantity. The formula for Laspeyre's price index method is as follows,

$$\boxed{P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100}$$

Where,

p_1 = Current year prices

p_0 = Base year prices

q_0 = Quantities consumed in base year.

2. Paasche's Index Method

In Paasche's index method, prices are weighted by the quantity consumed in the current year, rather than considering quantity consumed in base year. The formula for Paasche's index method is,

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

Where,

p_1 = Current year prices

p_0 = Base year prices

q_1 = Quantities consumed in current year.

3. Marshall Edgeworth's Index Method

In Marshall-Edgeworth's index method, both the current year as well as base year quantities are considered to calculate the index. The formula for Marshall-Edgeworth's Index method is as follows,

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

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

4. Fisher's Ideal Index Method

A statistician named "Fisher" introduced this method, which is a geometric mean of Laspeyre's and Paasche's methods. The formula used for this method is,

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

5. Dorbish and Bowley's Index Method

In this method, price index is the simple arithmetic mean of Laaspeyre's and Paasche's price indices. The formula for calculating the price index using Dorbish and Bowley method is as follows,

$$P_{01} = \frac{1}{2}\left[\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right] \times 100$$

6. Walsch's Index Method

In this method, price index is calculated by using the geometric mean of the base and current year quantities. The formula for Walsch's price index method is as follows,

$$P_{01} = \frac{\Sigma P_1 \sqrt{q_0 q_1}}{\Sigma P_0 \sqrt{q_0 q_1}} \times 100 \text{ or } P_{01} = \frac{\Sigma p_1 q}{\Sigma p_0 q},$$

where, $q = \sqrt{q_0 q_1}$

7. Kelly's Index Method

In this method, fixed weights are used to calculate the price index. It is also known as fixed weight aggregate method. The formula for Kelly's price index method is as follows,

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

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

Q16. Discuss the advantages and disadvantages of the following,

(a) Laspeyre's index method

(b) Paasche's index method

(c) Marshall-Edgeworth's index method

(d) Fisher's ideal index method.

Answer :

Advantages of Laspeyre's Index Method

The advantages of Laspeyre's index method are as follows,

1. It facilitates direct comparison of index of various periods with other periods.
2. It is not required to maintain a record of the consumed quantities in each period as this method takes into consideration only single quantity measure on the basis of the base period.

Disadvantage of Laspeyre's Index Method

The disadvantages of Laspeyre's index method are as follows,

1. It is upward biased method and thus overestimate price levels and inflation.
2. It considers only base year quantities for calculating price index.

Advantages of Paasche's Index Method

The advantages of Paasche's index method are as follows,

1. It integrates effects of changes in price and quantity consumed during the current year.
2. It gives a more effective estimation of changes of commodities or items when compared to the Laspeyre's method.

Disadvantages of Paasche's Index Method

The disadvantages of Paasche's index method are as follows,

1. It is costly and time consuming in nature as it needs to maintain the record of quantities consumed in each year.
2. It is required to recompute previous years index number is needed in order to show the effect of the new quantity weights.

Advantages of Marshall Edgeworth's Index Method

The advantages of Marshall-Edgeworth's index method are as follows,

1. It is easy to understand and calculate.
2. It satisfies both the unit and time reversal tests of consistency.
3. It uses current and base year's quantities and prices.

Disadvantages of Marshall Edgeworth's Index Method

The disadvantages of Marshall-Edgeworth's index method are as follows, index numbers.

2. It does not satisfy circular and factor reversal test of consistency.

Advantages of Fisher's Ideal Index Method

The advantages of Fisher's ideal index method are as follows,

1. It takes into consideration both the base year and the current year quantities.

2. It satisfies both time reversal test and Factor reversal test.

3. It is based on geometric mean which facilitates calculation of best index number

Disadvantages of Fisher's Ideal Index Method

Some of the disadvantages of Fisher's ideal index method are as follows,

1. It requires complicated computation which is lengthy in nature.

2. It not suitable for general use.

2.3.1 Quantity Index Numbers

Q17. Write about quantity index numbers. What are the various methods used for constructing quantity index numbers?

Answer :

Quantity Index Numbers

Quantity Index numbers measures the variation in the volume of goods manufactured or consumed or distributed in quantitative terms. It measures the change in quantitative terms like, increase or decrease in volume of production construction or employment.

While constructing quantity index numbers, current year production or sales data is compared with the base year data. In order to measure the changes in quantities, the values or prices are taken as weights. The formulae for quantity index numbers can be obtained from the formulae used in index numbers by taking (p) in place of (q) and taking (q) in place of (p).

Methods for Constructing Quantity Index Numbers

The following are the three methods which are used for constructing quantity index numbers,

(i) Laspeyre's Quantity Index Method

$$Q_{01} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times 100$$

Where,

q_1 = Quantity consumed in the current year.

q_0 = Quantity consumed in the base year.

p_0 = Base year prices.

(ii) Paasche's Quantity Index Method

$$Q_{01} = \frac{\sum q_1 p_1}{\sum q_0 p_1} \times 100$$

Where,

q_1 = Quantity consumed in the current year.

q_0 = Quantity consumed in the base year.

p_1 = Current year prices.

(iii) Fisher Quantity Index Method

$$Q_{01} = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}} \times 100$$

PROBLEMS ON INDEX NUMBERS

Q18. From the following data, calculate price index number by,

(i) Laspeyre's method

(ii) Paasche's method

(iii) Fisher's ideal method

(iv) Kelly's method

(v) Walsch's method

(vi) Marshall Edgeworth's method

(vii) Dorbish and Bowleys method.

Commodity	2005		2009	
	Price ₹	Qty	Price ₹	Qty
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

(March/April-12, Q13(b), (OU)

| March/April-11, Q13(b), (OU))

Solution :

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1	q_0q_1	$\sqrt{q_0q_1}$	$p_1\cdot\sqrt{q_0q_1}$	$p_0\cdot\sqrt{q_0q_1}$	$q=\frac{q_0+q_1}{2}$	p_0q	p_1q
A	20	8	40	6	320	160	240	120	48	6.93	277.2	138.6	7	140	280
B	50	10	60	5	600	500	300	250	50	7.07	424.2	353.5	7.5	375	450
C	40	15	50	15	750	600	750	600	225	15	750	600	15	600	750
D	20	20	20	25	400	400	500	500	500	22.36	447.2	447.2	22.5	450	450
Total					Σp_1q_0 = 2070	Σp_0q_0 = 1660	Σp_1q_1 = 1790	Σp_0q_1 = 1470			$\Sigma p_1\sqrt{q_0q_1}$ = 1898.6	$\Sigma p_0\sqrt{q_0q_1}$ = 1539.3		Σp_0q = 1565	Σp_1q = 1930

(i) Laspeyre's Index Method

$$P_{01} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100$$

$$= \frac{2070}{1660} \times 100$$

$$= 124.7$$

(ii) Paasche's Index Method

$$P_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100$$

$$= \frac{1790}{1470} \times 100$$

$$= 121.8$$

(iii) Fisher's Ideal Index Method

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

$$= \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$$

$$= \sqrt{1.247 \times 1.218} \times 100$$

$$= 123.24$$

(iv) Kelly's Method

$$P_{01} = \frac{\Sigma p_1q}{\Sigma p_0q} \times 100$$

$$= \frac{1930}{1565} \times 100$$

$$= 123.32$$

(v) Walsch's Method

$$P_{01} = \frac{\Sigma p_1\sqrt{q_0q_1}}{\Sigma p_0\sqrt{q_0q_1}} \times 100$$

$$= \frac{1898.6}{1539.3} \times 100$$

$$= 123.34$$

(vi) Marshall Edgeworths Method

$$P_{01} = \left[\frac{\Sigma p_1q_0 + \Sigma p_1q_1}{\Sigma p_0q_0 + \Sigma p_0q_1}\right] \times 100$$

$$= \left[\frac{2070 + 1790}{1660 + 1470}\right] \times 100$$

$$= \left[\frac{3860}{3130}\right] \times 100$$

$$= 123.32$$

(vii) Dorbish and Bowley's Method

$$P_{01} = \frac{1}{2}\left[\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} + \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right] \times 100$$

$$= \frac{1}{2}\left[\frac{2070}{1660} + \frac{1790}{1470}\right] \times 100$$

$$= \frac{1}{2}[1.247 + 1.218] \times 100$$

$$= \frac{1}{2}[2.465] \times 100$$

$$= 123.25$$

Q19. Calculate Quantity Index by (i) Laspeyre's Method (ii) Pasche's Method and (iii) Bowley' Method.

Commodity	Base Year		Current Year	
	Quantity	Rate	Quantity	Rate
Bread	10	3	8	3.25
Meat	20	15	15	20
Tea	2	25	3	23

Solution : Sept./Oct.-21, Q11 (OU)

Commodity	p_0	q_0	p_1	q_1	$p_1 q_0$	$p_0 q_0$	$p_1 q_1$	$P_0 q_1$
Bread	3	10	3.25	8	32.5	30	26	24
Meat	15	20	20	15	400	300	300	225
Tea	25	2	23	3	46	50	69	75
					$\Sigma p_1 q_0 = 478.5$	$\Sigma p_0 q_0 = 380$	$\Sigma p_1 q_1 = 395$	$\Sigma P_0 q_1 = 324$

(i) Laspeyre's Method

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$\therefore \quad P_{01} = \frac{478.5}{380} \times 100 = 125.92$$

(ii) Pasche's Method

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

$$P_{01} = \frac{395}{324} \times 100$$

$$\therefore \quad P_{01} = 121.91$$

(iii) Bowley's Method

$$P_{01} = \frac{1}{2} \times \left[\left(\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}\right) + \left(\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right)\right] \times 100$$

$$P_{01} = \frac{1}{2} \times \left[\frac{478.5}{380} + \frac{395}{324}\right] \times 100$$

$$P_{01} = \frac{1}{2} \times [1.2592 + 1.2191] \times 100$$

$$P_{01} = \frac{1}{2} \times [2.4783] \times 100$$

$$P_{01} = \frac{1}{2} \times 247.83$$

$$\therefore \quad P_{01} = 123.915.$$

Q20. 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

Solution : Jan.-21, Q12 (OU)

Note: As price and expenditure are given for base year and current year, divide expenditure of each commodity with their respective price to obtain quantity of base year (i.e., q_0) and quantity of current year (i.e., q_1).

Item	Base year Price (p_0)	Base Year Expenditure (p_0q_0)	Current Year Price (p_1)	Current year Expenditure (p_1q_1)	$q_0 = \frac{p_0q_0}{p_0}$	$q_1 = \frac{p_1q_1}{p_1}$	p_1q_0	p_0q_1
A	5	50	8	40	10	5	80	25
B	7	25	12	30	3.6	2.5	43.2	17.5
C	9	10	15	25	1.1	1.7	16.5	15.3
D	12	5	20	18	0.42	0.9	8.4	10.8
		90		**113**			**148.1**	**68.6**

(i) Laspeyre Method

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{148.1}{90} \times 100$$

$$= 164.55$$

(ii) Paasche Method

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

$$= \frac{113}{68.6} \times 100$$

$$= 164.72$$

(iii) Marshall - Edgeworth Methods

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

$$= \frac{148.1 + 113}{90 + 68.6} \times 100$$

$$= \frac{261.1}{158.6} \times 100$$

$$= 164.63.$$

Q21. Calculate Fishers ideal index from the following data.

Commodity	Base year 2015		Current year 2016	
	Price	Quantity (₹)	Price	Quantity (₹)
A	100	13	140	15
B	150	18	220	24
C	120	15	150	18
D	80	11	100	15
E	60	19	90	19

Solution :

July/Aug.-21, Q3 (MGU)

Calculation of Fisher's Ideal Index Number

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	100	13	140	15	1,820	1,300	2,100	1,500
B	150	18	220	24	3,960	2,700	5,280	3,600
C	120	15	150	18	2,250	1,800	2,700	2,160
D	80	11	100	15	1,100	880	1,500	1,200
E	60	19	90	19	1,710	1,140	1,710	1,140
					Σp_1q_0 =10,840	Σp_0q_0 =7,820	Σp_1q_1 =13,290	$\Sigma p0q_1$ =9,600

$$\text{Fisher's Ideal Index, } p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{10,840}{7,820} \times \frac{13,290}{9,600}} \times 100$$

$$= \sqrt{1.386 \times 1.384} \times 100$$

$$= 138.50$$

Q22. Calculate Fisher's Ideal Index from the following data.

Article	Base Year		Current Year	
	Quantity	Value	Quantity	Value
A	5	50	4	48
B	8	48	7	49
C	6	18	5	20

Solution :

Sept./Oct.-21, Q12 (OU)

In the above problem price of each article in not given, so to calculate the amount of price, divide total value by quantity.

i.e, $\frac{\text{Total Value}}{\text{Quantity}} = \text{Price}$

Article	Base Year			Current Year		
	Total Value	Price (p_0)	Quantity (q_0)	Total Value	Price (p_1)	Quantity (q_1)
A	50	10	5	48	12	4
B	48	6	8	49	7	7
C	18	3	6	20	4	5

Calculation of Fisher's Ideal Index

Article	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	10	5	12	4	60	50	48	40
B	6	8	7	7	56	48	49	42
C	3	6	4	5	24	18	20	15
					$\Sigma p_1q_0 = 140$	$\Sigma p_0q_0 = 116$	$\Sigma p_1q_1 = 117$	$\Sigma p_0q_1 = 97$

Fisher's Ideal Index, $p_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$

$$P_{01} = \sqrt{\frac{140}{116} \times \frac{117}{97}} \times 100$$

$$P_{01} = \sqrt{1.21 \times 1.21} \times 100$$

$$P_{01} = \sqrt{1.46} \times 100$$

$$P_{01} = 1.21 \times 100$$

$$\therefore \quad P_{01} = 121$$

Q23. Calculate Fisher's idle index from the following data.

Goods	A	B	C	D	E
P_0(₹)	18	14	16	10	12
Total Cost	1000	600	480	840	720
P_1(₹)	18	16	14	18	20
Total Cost	2,400	960	1050	900	800

Solution : May/June-19, Q7(a) (MGU

In the given problem, value of price and total cost is given. Thus, first we need to find out the quantity of the base year (q_0) and current year (q_1) by using the following formula,

$$\text{Quantity} = \frac{\text{Total Cost}}{\text{Price}}$$

The values of 'q_0' and 'q_1' are calculated as,

Goods	q_0	q_1
A	$\frac{1000}{18} = 55.56$	$\frac{2400}{18} = 133.33$
B	$\frac{600}{14} = 42.86$	$\frac{960}{16} = 60$
C	$\frac{480}{16} = 30$	$\frac{1050}{14} = 75$
D	$\frac{840}{10} = 84$	$\frac{900}{18} = 50$
E	$\frac{720}{12} = 60$	$\frac{800}{20} = 40$

Calculation of Fisher's Ideal Index

Goods	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	18	55.56	18	133.33	1000.08	1000.08	2399.94	2399.94
B	14	42.86	16	60	685.76	600.04	960	840
C	16	30	14	75	420	480	1050	1200
D	10	84	18	50	1512	840	900	500
E	12	60	20	40	1200	720	800	480
					Σp_1q_0 = **4817.84**	Σp_0q_0 = **3640.12**	Σp_1q_1 = **6109.94**	Σp_0q_1 = **5419.94**

Fisher's ideal price index,

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

$$= \sqrt{\frac{4,817.84}{3,640.12} \times \frac{6,109.94}{5,419.94}} \times 100$$

$$= \sqrt{1.3235 \times 1.1273} \times 100$$

$$= \sqrt{1.492} \times 100$$

$$= 1.2215 \times 100$$

$$= 122.15$$

2.4 TESTS OF CONSISTENCY OF INDEX NUMBERS – UNIT TEST, TIME REVERSAL TEST, FACTOR REVERSAL TEST AND CIRCULAR TEST

Q24. Explain various tests of consistency of index numbers.

OR

Discuss the following,

(i) Unit test

(ii) Time reversal test.

(iii) Factor reversal test

(iv) Circular test.

Answer :

Price index numbers have various weighted and unweighted index methods. Therefore, it is difficult to select a suitable method while constructing an index number. In order to overcome this issue, statisticians have suggested few tests for testing the adequacy or consistency of an index number. These tests include,

(i) Unit test

(ii) Time reversal test.

(iii) Factor reversal test

(iv) Circular test.

(i) Unit Test

According to unit test, the formula of index number should be independent of the units under which prices and quantities are quoted. All formulae satisfies this test except simple aggregative test.

(ii) Time Reversal Test

Time reversal test is basically used at checking whether the selected method would work for both forward and backward or not. According to this test, the formula should give exact ratio when compared with one point with the another i.e., for example,

$$P_{01} = \frac{1}{P_{01}} \text{ or } P_{01} \times P_{10} = 1$$

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

Only two methods, Laspeyre's and Paasche's do not satisfy the time reversal test. Besides these two, the other methods of index numbers satisfies the time reversal method.

Fisher's index method satisfies time reversal test.

Proof

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

If P_{10} is calculated, then the formula would be,

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

Time reversal test : $P_{01} \times P_{10} = 1$

$$\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = 1$$

Time reversal test is satisfied by Fisher's index method. Hence proved.

(iii) Factor Reversal Test

According to factor reversal test, when change in price is multiplied with change in quantity, it should give total change in value. It means, if the price of a commodity is increased by 3 times and its quantity has also increased by 4 times, then the total change in value would be 12 times than that of the former value. Thus, the formula for commodity whole price and quantity is p_0 and p_1 and q_0 and q_1 for base and current year is,

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

Except Fisher's ideal index, no other method satisfies the factor reversal test.

Proof

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

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

Factor reversal test : $P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

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

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

Fisher's ideal index satisfies the factor reversal test. Hence proved.

(iv) Circular Test

According to circular test, if P_{ab} is the price index whose base year is 'a' and current year is 'b', P_{bc} is the price index with base year 'b' and current year is 'c' and P_{ca} is the price index whose base year is 'c' and current year is 'a' then,

$$P_{ab} \times P_{bc} \times P_{ca} = 1$$

The following methods satisfies the circular test,

(a) Simple geometric mean of price relatives

(b) Simple aggregative index

(c) Weighted aggregative index.

Apart from these methods, no other method of price index satisfies the circular test.

PROBLEMS ON TESTS OF CONSISTENCY OF INDEX NUMBERS

Q25. From the following data, construct Fisher's Ideal Index Number and test whether it satisfies:

(a) Time Reversal Test and

(b) Factor Reversal Test

Commodity	Base Year		Current Year	
	Quantity	Price	Quantity	Price
A	21	11	27	13
B	26	8	33	6
C	26	6	21	10
D	16	7	26	5

Solution :

July/Aug.-21, Q5 (KU)

Calculation of Fisher's Ideal Index Number

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	11	21	13	27	273	231	351	297
B	8	26	6	33	156	208	198	264
C	6	26	10	21	260	156	210	126
D	7	16	5	16	80	112	80	112
					$\Sigma p_1q_0 = 769$	$\Sigma p_0q_0 = 707$	$\Sigma p_1q_1 = 839$	$\Sigma p_0q_1 = 799$

Fisher's Ideal Index, $p_{01} = \sqrt{\dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$$= \sqrt{\frac{769}{707} \times \frac{839}{799}} \times 100$$

$$= \sqrt{1.0877 \times 1.0501} \times 100$$

$$= 106.87$$

(a) Time Reversal Test

Time Reversal Test is satisfied when,

$p_{01} \times p_{10}$

$$p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} = \sqrt{\frac{769}{707} \times \frac{839}{799}} = \sqrt{1}$$

$$p_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{\frac{799}{839} \times \frac{707}{769}} = \sqrt{1}$$

$$p_{01} \times p_{10} = \sqrt{\frac{769}{707} \times \frac{839}{799} \times \frac{799}{839} \times \frac{707}{769}}$$

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

$$\boxed{p_{01} \times p_{10} = 1}$$

∴ This index number satisfies the time reversal test.

(b) Factor Reversal Test

Factor reversal test is satisfied when,

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

$$p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} = \sqrt{\frac{769}{707} \times \frac{839}{799}}$$

$$q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{799}{707} \times \frac{839}{769}}$$

$$\therefore \quad p_{01} \times q_{01} = \sqrt{\frac{769}{707} \times \frac{839}{799} \times \frac{799}{707} \times \frac{839}{769}}$$

$$= \sqrt{\left(\frac{839}{707}\right)^2}$$

$$= \frac{839}{707} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \text{ or } \frac{839}{707} = \frac{839}{707}$$

$$\boxed{\therefore p_{01} \times q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}}$$

∴ This index number satisfies the factor reversal test.

Q26. Calculate Fisher's ideal index from the following data and show how it satisfies T.R.T and F.R.T.

Commodities	2010		2012	
	Price	Total Value	Price	Total Value
P	8	80	10	110
Q	10	90	12	108
R	16	256	20	340

Solution : Oct./Nov.-14, Q13(b) (OU)

In the given problem, quantity of each commodity is not given. Therefore, calculate amount of quantity by using following formula,

$$\text{Quantity} = \frac{\text{Total value}}{\text{Price}}$$

Commodity	2010			2012		
	Total value	Price (p_0)	Quantity (q_0)	Total value	Price (p_1)	Quantity (q_1)
P	80	8	10	110	10	11
Q	90	10	9	108	12	9
R	256	16	16	340	20	17

Calculation of Fisher's Ideal Index

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
P	8	10	10	11	100	80	110	88
Q	10	9	12	9	108	90	108	90
R	16	16	20	17	320	256	340	272
					$p_1q_0 = 528$	$p_0q_0 = 426$	$p_1q_1 = 558$	$p_0q_1 = 450$

Fisher's Ideal Index,

$$p_0 = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{528}{426} \times \frac{558}{450}} \times 100$$

$$= \sqrt{1.24 \times 1.24} \times 100 = \sqrt{1.5376} \times 100$$

$$= 124$$

Time Reversal Test

Time Reversal Test is satisfied when,

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

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

$$p_{01} \times p_{10} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}}$$

$$p_{01} \times p_{10} = \sqrt{\frac{528}{426} \times \frac{558}{450} \times \frac{450}{558} \times \frac{426}{528}}$$

$$= 1$$

Hence, Time Reversal Test is satisfied.

Factor Reversal Test

Factor Reversal Test is satisfied when,

$$p_{01} \times q_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

$$q_{01} = \sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_0p_1}}$$

$$p_{01} \times q_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_0p_1}}$$

$$p_{01} \times q_{01} = \sqrt{\frac{528}{426} \times \frac{558}{450} \times \frac{450}{426} \times \frac{558}{528}}$$

$$p_{01} \times q_{01} = 1.31$$

$$\frac{\Sigma p_1q_1}{\Sigma p_0q_0} = \frac{558}{426} = 1.31$$

$$\therefore \; p_{01} \times q_{10} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

Hence, Factor Reversal Test is satisfied.

2.5 BASE SHIFTING

Q27. What is Base Shifting? Explain it with an illustration.

Answer :

Base Shifting

Base Shifting refers to the process of shifting a base period of an index. It is also known as changing of base. It helps in calculating the index numbers based on new base. As the old base gets outdated, it is required to shift or change such old base to new-base.

The base shifting can be calculated by using the following formula,

$$\text{Base Shifting} = \frac{\text{Current year's old index number}}{\text{New base year's old index number}} \times 100$$

Illustration

From the following information of index numbers calculate base shifting.

Consider 2001 base year of old index number and 2004 as new base year.

Years	2001	2002	2003	2004	2005
Index Numbers	100	120	150	180	225

Solution :

The calculation of index number based on new base year is,

$$= \frac{\text{Current year's old index number}}{\text{New base year's old index number}} \times 100$$

Years	Old Index Numbers (Base Year 2001 = 100)	New Index Numbers (Base Year 2004) (2004 = 180)
2001	100	$\frac{100}{180} \times 100 = 55.55$
2002	120	$\frac{120}{180} \times 100 = 66.67$
2003	150	$\frac{150}{180} \times 100 = 83.33$
2004	180	$\frac{180}{180} \times 100 = 100$
2005	225	$\frac{225}{180} \times 100 = 125$

PROBLEMS ON BASE SHIFTING

Q28. The index of 2010 is 100. It rises by 4% 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.

Solution :

Jan.-21, Q11 (OU)

Calculation of Index Numbers for Base Year 2010 and Change of Base Year 2014

Years	Old Index Number (Base Year 2010 = 100) $= \frac{100 \pm \%}{100} \times$ 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	$\frac{100+4\%}{100} \times 100 = 104$	$\frac{100}{116.59} \times 104 = 89.20$
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	$\frac{100+10\%}{100} \times 105.99 = 116.59$	$\frac{100}{116.59} \times 116.59 = 99.99$
2015	$\frac{100-3\%}{100} \times 116.59 = 113.09$	$\frac{100}{116.59} \times 113.09 = 96.99$
2016	$\frac{100+8\%}{100} \times 113.09 = 122.14$	$\frac{100}{116.59} \times 122.14 = 104.76$

Q29. 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.

Solution : May/June-18, Q10(a) (OU)

The calculation of index number based on new base year is $= \frac{\text{Current year's old index number}}{\text{New base year's old index number}} \times 100$

Years	Old Index Numbers (Base year 2007 = 100)	New Index Numbers (Base year 2012 = 120)
2007	100	$\frac{100}{120} \times 100 = 83.33$
2008	120	$\frac{120}{120} \times 100 = 100$
2009	122	$\frac{122}{120} \times 100 = 101.67$
2010	116	$\frac{116}{120} \times 100 = 96.67$
2011	120	$\frac{120}{120} \times 100 = 100$
2012	120	$\frac{120}{120} \times 100 = 100$
2013	137	$\frac{137}{120} \times 100 = 114.17$
2014	136	$\frac{136}{120} \times 100 = 113.33$
2015	149	$\frac{149}{120} \times 100 = 124.17$
2016	156	$\frac{156}{120} \times 100 = 130$
2017	137	$\frac{137}{120} \times 100 = 114.17$

2.6 SPLICING OF INDEX

Q30. What is Splicing? What are its conditions, importance and types?

Answer :

Splicing

Splicing refers to the process of combining two or more overlapping indices in which various bases are converted into a single series. In other words, splicing of index numbers can be defined as the procedure of converting two or more series of index numbers of different bases into a single or continuous series of index numbers.

Conditions of Splicing

The following conditions are essential for splicing the index numbers,

1. It should include atleast two series of index numbers should be from the same group.

2. It must include base period of different indices which are built on different series.

3. It includes indices of two bases which should be there in a year. For example, in 2015 there may be two index numbers with 2005 and another with 2013.

Importance of Splicing

The importance of splicing of index numbers is highlighted in the following points,

1. It is important in the situation where a series of index numbers with old base is discontinued and new series of index numbers is constructed with new base year.

2. It helps to bring two or more disjoint index series under a common base.

3. It helps to maintain continuous series in index numbers.

4. It helps to compare the index numbers of various years.

Types of Splicing

Generally, splicing is categorized into two types. They are as follows,

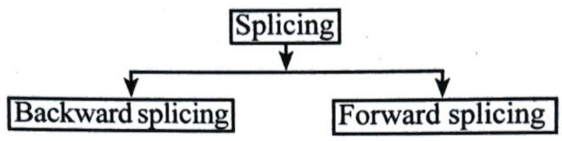


Figure: Types of Splicing

1. Backward Splicing

Backward splicing is used for splicing a new series of indices to continue with old series of indices. The formula used for backward splicing is as follows,

$$\text{Backward Splicing} = \frac{\text{Index 'A' of Current Year}}{\text{Index 'A' of Common Year}} \times 100$$

2. Forward Splicing

Forward splicing is used for splicing old series of indices to continue with new series of indices. The formula used for forward splicing is as follows,

$$\text{Forward Splicing} = \frac{\text{Index 'B' of Current Year} \times \text{Index 'A' of Common Year}}{100}$$

PROBLEMS ON SPLICING OF INDEX

Q31. The following table gives two series of index numbers with 2005 and 2010 as base year. Obtain a continuous series of index number by considering 2010 as base year.

Years	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Index No. (2005 = 200)	200	240	250	280	300	320				
Index No. (2010 = 200)						200	220	240	260	280

Solution :

$$\text{Backward Splicing} = \frac{\text{Index A of Current Year}}{\text{Index A of Common Year}} \times 100$$

Years	Index Number (2005 = 200)	Index Number (2010 = 200)	Backward Splicing
2005	200		$\frac{200}{320} \times 100 = 62.5$
2006	240		$\frac{240}{320} \times 100 = 75$

2007	250		$\frac{250}{320}\times 100 = 78.125$
2008	280		$\frac{280}{320}\times 100 = 87.50$
2009	300		$\frac{300}{320}\times 100 = 93.75$
2010	320	200	$\frac{320}{320}\times 100 = 100$
2011		220	220
2012		240	240
2013		260	260
2014		280	280

Q32. The following table gives two series of index numbers with 2006/2010 as base year. Obtain a combined series of index numbers with 2006 as base year.

Years	2006	2007	2008	2009	2010	2011	2012	2013	2014
Index No. (2006 = 100)	100	105	118	120	125				
Index No. (2010 = 100)					100	120	132	145	160

Solution :

$$\text{Forward Splicing} = \frac{\text{Index B of Current Year} \times \text{Index A of Common Year}}{100}$$

Years	Index No. (2006 = 100) Series A (Forward Splicing)	Index No. (2010 = 100) Series B
2006	100	
2007	105	
2008	118	
2009	120	
2010	125	100
2011	$\frac{120\times 125}{100} = 150$	120
2012	$\frac{132\times 125}{100} = 165$	132
2013	$\frac{145\times 125}{100} = 181.25$	145
2014	$\frac{160\times 125}{100} = 200$	160

2.7 DEFLATING OF INDEX

Q33. What do you mean by deflating? Explain with an example.

Answer :

Deflating

Deflating refers to the process of making allowances for the impact of changing prices. A rise in price level means a reduction in the purchasing power of money.

Example

The price of rice in 2010 was ₹ 28 per kg but in 2013 the price increased to ₹ 45 per kg. It means a person who could buy 1 kg rice for ₹ 28 in 2010 would be able to buy only half kg rice in 2013 as their income level remains same. During the period of inflation, the purchasing power of money is the reciprocal of the price index. This reciprocal relationship can be shown in the form of formula as follows,

$$\text{Purchasing power of money} = \frac{1}{\text{Price index}}$$

Suppose, if price of a commodity is 30%, price index is 1.30 and the purchasing power of rupee is $1/1.30 = 0.77$ or 77 paise. As value of money decreases with the increase in prices wage workers or salaried people show more interest in real wage rather than money wage. Real waged can be obtained by using the following formulae,

$$\text{Real wage} = \frac{\text{Money wage}}{\text{price index}} \times 100$$

$$\text{Real wage or income index No.} = \frac{\text{Index of money wages}}{\text{Consumer price index}}$$

PROBLEMS ON DEFLATING OF INDEX

Q34. The annual income of a employee and the general index numbers of price during 2011-2019 is given in the following table. Prepare index number to show the changes in real income of the employee and comment on price increase,

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
Income	3,500	4,000	5,300	5,500	6,400	6,000	7,200	7,500	6,800
Price Index No.	100	120	145	160	250	320	450	530	600

Solution :

Calculation of index Numbers Showing Changes in the Real income of the Employee

Years (1)	Income (₹) (2)	Price Index No. (3)	Real Income (4) = [2÷3]× 100	Real income Index No. (5)
2011	3,500	100	$\frac{3,500}{100} \times 100 = 3,500$	3,500/3,500 × 100=100
2012	4,000	120	$\frac{4,000}{120} \times 100 = 3,333.33$	3,333.33,500 ×100 = 95.23
2013	5,300	145	$\frac{5,300}{145} \times 100 = 3,655.17$	3,655.17/3,500 ×100 = 104.43
2014	5,500	160	$\frac{5,500}{160} \times 100 = 3,437.5$	3,437.5/3,500 × 100 = 98.2
2015	6,400	250	$\frac{6,400}{250} \times 100 = 2,560$	2,560 / 3,500 × 100 =73.14
2016	6,000	320	$\frac{6000}{320} \times 100 = 1,875$	1,875/3,500 × 100 =53.57
2017	7,200	450	$\frac{7,200}{450} \times 100 = 1,600$	1,600/3,500 ×100 = 45.71
2018	7,500	530	$\frac{7,500}{530} \times 100 = 1,415.09$	1,415.09/3,500 × 100 = 40.43
2019	6,800	600	$\frac{6,800}{600} \times 100 = 1,133.3$	1,133.3/3,500 × 100 = 32.38

2.8 CONSUMER PRICE INDEX

Q35. What are consumer price index numbers? Explain the steps in their construction and list out their uses.

Answer :

Consumer Price Index Numbers

Consumer price index numbers also known as cost of living index numbers. They are used to measure the purchasing power of a particular class of people in relation to the changes in retail their prices. In other words, it studies how price variations effect the cost of living or purchasing power of a group of people.

While constructing cost of living index number, a particular section of society is selected like rich, middle, poor and a study is conducted to know how price variations effect the consumption levels of that section. Based on such information, Consumer Price Index Numbers (CPIN) is constructed.

Steps in Construction of Consumer Price Index Numbers (CPIN)

The steps in construction of consumer price index numbers are as follows,

Step-1: Selection of Group of People

A group of people or class of people is selected to construct CPIN. Apart from class of people, the area (i.e., rural or urban, city or town) should be clearly specified. The group of people selected for constructing CPIN must be homogenous to a maximum extent.

Step-2: Conducting Family Budget Enquiry

An enquiry of family budget is conducted to know how much money an average family spends on the consumption of different items. These items are broadly categorized into five groups namely,

(a) Food

(b) Clothing

(c) Fuel and lighting

(d) House Rent and

(e) Miscellaneous.

Each of the above groups is further sub- categorised into small groups, Example, The group "food" is subdivided into cereals, like wheat, rice, pulses meat, fish , milk, fruits, vegetables and so on.

Step-3: Price Quotations

While gathering information about retail prices a proper care should be taken as retail prices varies from place to place and from shop to shop. Information about retail prices should be gathered from those local markets where selected class of people are located.

Uses of CPIN

The uses of CPIN are as follows,

1. They are used in the preparation of wage contracts and wage negotiations.
2. They assists the government and business organizations in deciding Dearness Allowance [D.A] to be paid to their employees.
3. They are used for deflating income and value series in National Income of the country.
4. They are used to measure purchasing power of money.
5. They assist in calculating real wage by considering the variations in money income and price level.

Q36. Explain the methods in construction of consumer price index.

Answer :

Methods of Construction of Consumer Price Index Numbers

It was observed that the significance of consumption items is different for different groups of people. Even the people belonging to same class might have different opinion regarding the significance of consumption items.

This is the reason why cost of living index is determined as weighted indices by taking into account the relative significance of consumption items. The significance of consumption items is determined based upon the money spent by people on various items. Further cost of living index numbers are constructed by adopting the following two methods,

1. Aggregate Expenditure Method/Weighted Aggregate Method

Under this method the quantities (q) consumed in the base year are taken as weights. It can be expressed as,

$$\text{Consumer Price Index} = \frac{\text{Total expenditure in current year}}{\text{Total expenditure in base year}}$$

$$= \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

Where,

p_1 = Current year price

p_0 = Base year price

q_0 = Base year quantity.

2. Family Budget Method/Method of Weighted Relatives

Under this method, weighted average of price relatives are calculated to obtain cost of living index. Where, weights are equal to the quantities consumed in the base year. It can be expressed as,

$$\text{Consumer price index} = \frac{\Sigma WI}{\Sigma W}$$

Where, I = Price relative $= \frac{p_1}{p_0} \times 100$ and $w = p_0 q_0$,

By substituting the values of W and I, the following is obtained,

$$\text{Consumer Price Index} = \frac{\sum p_0 q_0 \left[\frac{p_1}{p_0} \times 100\right]}{\sum p_0 q_0}$$

$$= \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

PROBLEMS ON CONSUMER PRICE INDEX

Q37. A textile worker in the city of Ahmedabad earns ₹ 850 per month. The cost of living index for January, 1988 is given as 160. Using the following data, find out the amounts he spends on

(i) Food and

(ii) Rent

Group	Expenditure (₹)	Group Index
Food	?	195
Clothing	130	192
Rent	?	145
Fuel & Lighting	105	123
Miscellaneous	80	106

Solution : July/Aug.-21, Q4 (MGU)

Let the expenditure on food and rent be ₹ x and ₹ y respectively.

Group	Expenditure (W)	Group Index (I)	(IW)
Food	x	195	195X
Clothing	130	192	24,960
Rent	y	145	145Y
Fuel & Lighting	105	123	12,915
Miscellaneous	80	106	8,480
	$\Sigma W = 315 + X + Y$		$\Sigma IW = 195X + 145Y + 46,355$

Total Earning is ₹ 850

$315 + x + y = 850$

$x + y = 850 - 315 = 535$ ....(i)

Cost of living index $= 160 = \frac{\Sigma IW}{\Sigma W}$

$$160 = \frac{195x + 145y + 46,355}{850}$$

$195x + 145y = 850 \times 160 - 46,355$

$195x + 145y = 1,36,000 - 46,355$

$195x + 145y = 89,645$ ....(ii)

Now, we have two equations (i) and (ii). By considering the least value of equation (ii) i.e. 145, multiplying it with equation (i) both sides, we get

$145x + 145y = 77575$ .... (iii)

Now, simplifying the equations (ii) and (iii)

$$195x + 145y = 89645$$
$$145x + 145y = 77575$$
$$50x + 0 = 12070$$

$x = \frac{12070}{50} = 241.4$

$\therefore x = 241.4$

Now, substituting the value of X in equation (i),

$x + y = 535$

$241.4 + y = 534$

$y = 535 - 241.4$

$\therefore y = 293.6$

Therefore, the amount spent on the food (x) is ₹ 241.4 and on rent (y) ₹ 293.6.

Q38. Calculate cost of living index from the following data:

Article	Price in Base year (₹)	Price in Current year (₹)	Quantity in Base year (₹)
A	6	8	50
B	2	3	100
C	5	6	60
D	10	12	30

May/June-18, Q3(b) (KU)

Solution :

Calculation of Cost of Living Index Numbers by Aggregate Expenditure Method

Article	Prices in (₹)			Aggregate Expenditure	
	Base year $[P_0]$	Current year $[P_1]$	Base Year q_0	P_1q_0	P_0q_0
A	6	8	50	400	300
B	2	3	100	300	200
C	5	6	60	360	300
D	10	12	30	360	300
				$\Sigma P_1q_0 = 1,420$	$\Sigma P_0q_0 = 1,100$

$$\text{Cost of Living Index Number} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{1,420}{1,100} \times 100$$

$$= 129.09$$

Calculation of Living Index Numbers by Family Budget Method

Article	Prices in (₹)		Quantity	Price Relative (1)	W = Value in Base (2)	
	Base year $[P_0]$	Current year $[P_1]$	Base year (q_0)	(I) $P = P_1/P_0 \times 100$	$W = P_0q_0$	IW
A	6	8	50	$8/6 \times 100 = 133.33$	300	39,999
B	2	3	100	$3/2 \times 100 = 150$	200	30,000
C	5	6	60	$6/5 \times 100 = 120$	300	36,000
D	10	12	30	$12/10 \times 100 = 120$	300	36,000
					$\Sigma W = 1,100$	$\Sigma IW = 1,41,999$

$$\text{Cost of Living Index Number} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{1,41,999}{1,100}$$

$$= 129.09$$

EXERCISE AND PRACTICE QUESTIONS

SHORT QUESTIONS

THEORY

Q1. Define Index Numbers. [Refer, Q1]

Q2. Importance of Index Numbers. [Refer, Q2] May/June-18, Q2 (OU)

OR

Advantages of Index Numbers March/April-17, Q8 (OU)

OR

Explain the uses of Index Numbers. March/April-14, Q7 (OU)

Q3. Types of Index Numbers [Refer, Q3] (May/June-19, Q2 (OU) | Oct./Nov.-14, Q4 (OU))

Q4. Marshall Edge Worth Method. [Refer, Q4] May/June-19, Q2 (MGU)

Q5. What is "Cost of living Index"? [Refer, Q5] May/June-18, Q1(b) (KU)

PROBLEMS

Q6. Calculate index number by simple aggregative method. [Refer Similar, Q6]

	A	B	C	D
Price in 2005 (₹)	162	256	257	132
Price in 2007 (₹)	171	164	189	145

(Ans: 82.0)

Q7. Calculate index number by simple aggregative method. [Refer Similar, Q6]

	A	B	C	D
Price in 2005 (₹)	324	512	514	264
Price in 2007 (₹)	342	328	378	290

(Ans: P_{01} = 165.8)

Q8. The prices for period 2006 to 2008 are given below:

Year	Price
2006	50
2007	52
2008	54

Calculate simple price index taking 2016 as base year. [Refer Similar, Q6]

(Ans: $P_{2006,\,2006}$ = 100%; $P_{2006,\,2007}$ = 104%; $P_{2006,\,2008}$ = 108%)

Q9. From the following data calculate a price index based on price relatives method using Arithmetic Mean. [Refer Similar, Q7]

Commodity	A	B	C	D	E	F
Price 2015 (₹)	90	120	40	100	170	240
Price 2016 (₹)	110	140	60	150	180	260

(Ans: P_{01} = 251.034)

Q10. Calculate Index number by Average Price Relative Method by using Arithmetic Mean. [Refer Similar, Q8]

Commodity	P	Q	R	S	T
Price 2019	4	12	20	10	24
Price 2020	8	16	30	10	16

(Ans: 260)

Q11. Calculate index number for the following data by arithmetic average of price relatives method. [Refer Similar, Q8]
March/April-16, Q7 (OU)

Commodities	A	B	C	D	E	F
Price in 1980	40	60	20	50	80	100
Price in 1990	50	60	30	70	90	110

(Ans: P_{01} = 122.916).

ESSAY QUESTIONS

THEORY

Q1. Define Index Number. What are its features and uses? [Refer, Q11] May/June-19, Q10(a) (OU)

OR

What is Index Number? Write importance of index numbers. May/June-18, Q3(a) (KU)

Q2. What is the importance and limitations of index numbers? Explain. [Refer, Q12]
May/June-19, Q7(b) (MGU)

Q3. Discuss the problems in the construction of Index Numbers. [Refer, Q14]

Q4. What is Splicing? What are its conditions, importance and types? [Refer, Q31]

Q5. What are consumer price index numbers? Explain the steps in their construction and list out their uses. [Refer, Q36]

PROBLEMS

Q6. Compute a price index from the following data,

(a) Simple aggregative method

(b) Average of price relative method by using Arithmetic mean and Geometric mean. [Refer Similar, Q7, Q9]

Commodity	A	B	C	D	E	F
Price in 2010 (₹)	10	20	30	40	50	25
Price in 2015 (₹)	15	25	30	40	55	35

(Ans: (a) P_{01} = 114.29 (b) 120.83).

Q7. From the following data, calculate Laspeyre's, Paache's, Dorbish and Bowley, Marshall Edgeworth methods and Fisher's Ideal Index. [Refer Similar, Q19]

Item	Base Year		Current Year	
	Price(₹)	Qty(KG)	Price(₹)	Qty(KG)
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

(Ans: 139.71; 139.88; 139.8; 139.79; 139.796).

Q8. Calculate Fisher's ideal index from the following data and show how it satisfies Time Reversal Tests. [Refer Similar, Q23]

Items	Price		Quantity	
	2010	2011	2010	2011
A	8	20	50	60
B	2	6	15	10
C	1	2	20	25
D	2	5	10	8
E	1	5	40	30

(Ans: 266.61).

Q9. Compute Fisher's Ideal Index number and show how its satisfies the Time Reversal and factor reversal tests. [Refer Similar, Q23]

Commodities	Base year 2011		Current year 2018	
	Price	Qty	Price	Qty
P	10	100	20	140
Q	8	150	8	200
R	12	120	18	160
S	20	80	30	80
T	16	160	24	200

(Ans: 148.84).

Q10. Calculate the index number using both the aggregate expenditure method and family budget method for the year 2013 with 2000 as the base year from the following data. [Refer Similar, Q29]

Commodity	Qty in Units 2000	Price P.U 2000 (₹)	Price P.U 2013 (₹)
A	100	8.00	12.00
B	25	6.00	7.50
C	10	5.00	5.25
D	20	48.00	52.00
E	25	15.00	16.50
F	30	9.00	27.00

(Ans: 142.1).

Q11. Compute price index and quantity index numbers for the year 2010 with 2005 as base year using,

(i) Laspeyre's method

(ii) Paasche's method. [Refer Similar, Q20]

March/April-13, Q13(a) (OU)

Commodity	Quantity (units)		Value (₹)	
	2005	2010	2005	2010
A	100	150	500	900
B	80	100	320	500
C	60	72	150	360
D	30	33	360	297

(Ans: (i) Laspeyre's Price Quantity Indices P_{01} = 118.04 , Q_{01} = 129.77 (ii) Paasche's Price and Quantity Indices P_{01} = 119.18 , Q_{01} = 131.02)

Q12. Calculate Fisher's Ideal Index number from the data: [Refer Similar, Q23]

Commodities	2005		2010	
	Price Per Unit (₹)	Total Expenditure (₹)	Price Per Unit (₹)	Total Expenditure (₹)
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

March/April-16, Q13(a) (OU)

(Ans: P_{01} = 112.8).

Q13. Compute Laspeyre's and Paasche's index numbers from the following data: [Refer Similar, Q19]

	Base Year		Current Year	
	Price (₹)	Expenditure	Price (₹)	Expenditure
P	2	40	5	75
Q	4	16	8	40
R	1	10	2	24
S	5	25	10	60

Sept./Oct.-15, Q13(b) (OU)

(Ans: Laspeyre's Index Method p_{01} = 221.98 ; Paasche's Index Method p_{01} = 216.30).

Q14. The following are the index numbers of a commodity taking 2011 as the base:

Years	2011	2012	2013	2014	2015
Index Numbers	200	240	300	360	450

Find the index numbers by changing the base to 2013. [Refer Similar, Q30]

(Ans: 66.67, 80, 100, 120, 150).

Q15. Compute Price Index Number by using:

(i) Paasches and

(ii) Marshal and Edgeworth methods. [Refer Similar, Q19] May/June-19, Q10(b) (OU)

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

(Ans: Paasche's Method = 121.7, Marshall Edgeworth's Method = 121.3)

Q16. From the following data calculate Price Index Number by using May/June-18, Q10(b) (OU)

(i) Paasche's Method and (ii) Marshall Edgeworth Method. [Refer Similar, Q19]

Item	Base Year		Current Year	
	Price (₹)	Expenditure (₹)	Price (₹)	Expenditure (₹)
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

(Ans: Paasche's Method = 147.29; Marshall Edgeworth's Method = 148.66)

Q17. Calculate Fisher's Ideal Index Number and test whether it satisfies Time Reversal and Factor Reversal Test for the following data. [Refer Similar, Q26] Oct./Nov.-16, Q13(b) (OU)

Commodity	Base year		Current year	
	Price (₹)	Qty (kg)	Price (₹)	Qty (kg)
A	32	50	30	50
B	30	35	25	40
C	16	55	18	50

(Ans: Time Reversal Test = 1; Factor Reversal Test = $\frac{3400}{3530}$)

INTERNAL ASSESSMENT/EXAM

I Multiple Choice

1. __________ is a numerical value characterizing the change in complex economic phenomena over a period of time or space. [ ]

   (a) Current value (b) Base value
   (c) Index number (d) Topic number

2. $\frac{\text{Current year}}{\text{Base year}} \times 100$ is the formula of __________. [ ]

   (a) Multiple Price index (b) Double Price Index
   (c) Single – Double Price Index (d) Single Price Index

3. Laspeyre's Index Method was introduced by __________. [ ]

   (a) Etiebne Laspeyres (b) Herman Laspeyres
   (c) Shane Laspeyres (d) Angelo Laspeyres

4. $P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$ is the formula of _______. [ ]

   (a) Paasche's Index Method (b) Laspeyre's Price Index Method
   (c) Fishers Ideal Index Method (d) Marshall - Edgeworth Index Method

5. In ________ index method, prices are weighted by the quantity consumed in the current year, rather than considering quantity consumed in base year. [ ]

   (a) Laspeyre's (b) Fishers
   (c) Paasche's (d) Marshall-Edgeworth's

6. In __________ index method, both the current year as well as base year quantities are considered as to calculate the index. [ ]

   (a) Marshall-Edge worth's (b) Fishers
   (c) Paasche's (d) Laspeyre's

7. Marshall-Edgeworth's index method was proposed by __________ mathematicians. [ ]

   (a) 3 (b) 2
   (c) 1 (d) 6

8. There are ______ tests of consistency of index numbers. [ ]

   (a) 1 (b) 2
   (c) 3 (d) 4

9. $\frac{\text{Current year's old index number}}{\text{New base year's old index number}} \times 100$ is the formula of _________. [ ]

   (a) Calculation of index number based on new base year
   (b) Calculation of index number based on old base year
   (c) Calculation of index number based on current base year
   (d) None of the above

10. ________ is used for splicing old series of indices to continue with new series of indices. [ ]

    (a) Base shifting (b) Forward splicing
    (c) Backward splicing (d) None of the above

II Fill in the Blanks

1. __________ are used to measure the changes that takes place in the total value of the variable.
2. __________ measures the variation in the volume of goods manufactured or consumed or distributed in quantitative terms.
3. __________ are used for measuring differences in the magnitude of a group of related variables.
4. Index numbers are __________ by which changes in the magnitude of a phenomenon are measured from time to time or place to place.
5. __________ test is basically used for checking whether the selected method would work both backward or forward or not.
6. According to __________ test, when change in price is multiplied with change in quantity, it should give total change in value.
7. Index numbers are classified into __________ types.
8. In value index numbers, value = __________.
9. Formula of Marshall-Edgeworth's index method is __________.
10. __________ refers to the process of shifting a base period of an index.

KEY

I. Multiple Choice

1. (c)
2. (d)
3. (a)
4. (b)
5. (c)
6. (a)
7. (b)
8. (d)
9. (a)
10. (b)

II. Fill in the Blanks

1. Value index numbers
2. Quantity index numbers
3. Index numbers
4. Series of numbers
5. Time reversal
6. Factor reversal
7. 3
8. Price × Quantity
9. $P_{01} = \dfrac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$
10. Base shifting.

III Very Short Questions and Answers

Q1. Define Index Numbers.

Answer :

According to Maslow, "Index number is a numerical value characterizing the change in the complex economic phenomena over a period of time or space".

According to Croxton and Cowden, "Index numbers are devices measuring differences in the magnitude of a group of related variables".

Q2. What is "Cost of Living Index"?

Answer :

Consumer price index numbers is also known as cost of living index number. It is used to measure the purchasing power of a particular class of people in relation to the changes in retail prices. In other words, it studies how price variations effect the cost of living or purchasing power of a group of people.

Q3. Write about Weighted Price Indexes.

Answer :

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.

Q4. Write about Quantity Index Numbers.

Answer :

Quantity Index numbers measures the variation in the volume of goods manufactured or consumed or distributed in quantitative terms. It measures the change in quantitative terms like, increase or decrease in volume of production construction or employment.

Q5. What is Splicing?

Answer :

Splicing refers to the process of combining two or more overlapping indices in which various bases are converted into a single series. In other words, splicing of index numbers can be defined as the procedure of converting two or more series of index numbers of different bases into a single or continuous series of index numbers.

UNIT 3

TIME SERIES

SYLLABUS

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

LEARNING OBJECTIVES

- ✓ Concept of Time Series Analysis.
- ✓ Components of Time Series.
- ✓ Methods such as Semi Averages, Moving Averages and Least Square Method.
- ✓ Concept of Deseasonalisation of Data.
- ✓ Uses and Limitations of Time Series.

INTRODUCTION

Time series is an arrangement of statistical data in a chronological order i.e., in accordance with the time of its occurrence. Secular trend or long term movements, short term fluctuations or periodic movements and random or erratic fluctuations are the components of time series. The analysis of time series is not only used by the economist and business but it is also followed and used by the scientist, astronomist, geologist, sociologist, biologist and researchers.

The various methods which are used for measuring trend component of Time Series are Semi-Average Method, Moving Average Method and Least Square Method. Deseasonalisation of data refers to the process of eliminating seasonal fluctuations from the given time series for, it facilitates in adjusting the given time series for seasonal fluctuations and therefore left out with variables like trend component, cyclical and irregular variations.

PART-A
SHORT QUESTIONS AND ANSWERS

Q1. Time Series Analysis

Answer : Sept/Oct-21, Q3(OU)

"A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables".

– Ya-Lun Chou

Time series refers to the arrangement of statistical data in chronological order i.e., according to the time of occurrence. It represents the changing moments of variables over a particular period of time.

Time series plays an important role in business and economics. Thus, Economists developed many statistical techniques for analyzing time series data. However, these techniques can also be applied to study time series of other disciplines which are not related to economics and statistics like natural sciences, social sciences etc.

The functional relationship of time series can be mathematically represented as,

$$y = f(t)$$

Where, y = variable under consideration

f = functional relationship

t = times $t_1, t_2, t_3, \ldots t_n$.

Q2. Utility of Time Series Analysis. May/June-19, Q4 (OU)

OR

What are the uses of time series? May/June-18, Q4 (OU)

Answer :

The utility or uses of time series analysis are as follows,

1. Understanding Past Behaviour

It helps in understanding the past behaviour by considering the changes that have taken place in the past. It predicts the future behaviour with the help of past data.

2. Planning Future Operations

It helps in future planning by forecasting the events and their relationship. If regular occurrence of any event is there over a long period then such event is considered to predict the future.

3. Evaluating Current Accomplishments

It helps in evaluating the performance by comparing the actual performance with the expected performance and analyze the reasons of variations.

4. Facilitates Comparison

It compares the different periods and various conclusions are drawn upon it. But it is not necessary that everyone should believe on it because statisticians are not foretellers, they cannot predict 100% accurate results for future events.

Q3. Objectives of Time Series

Answer : May/June-19, Q3 (MGU)

There are two main objectives of time series which are as follows,

1. Studying Past Behaviour of Data

One of the important objective of time series analysis is to identify the patterns in the historical data and isolate the effects of various forces or factors. This helps in predicting the value of variable and also helps in planning for future.

2. Reviewing and Evaluating a Plan

Time series data forms the basis for review and evaluation of a plan. For example, a company may uses time series analysis in its evaluation policy for controlling inflation which is carried out by studying various price indices.

Q4. What are seasonal variations? Write three features of seasonal variations.

Answer : May/June-18, Q1(c) (KU)

Seasonal variations refers to those periodic movements which occur regularly every year and have their origin in the nature of the year itself. They occur in regular and periodic manner over a span of less than a year, i.e., during a period of twelve months and have the same or almost the same pattern year after year. The causes of seasonal variations are climate in its widest sense (natural causes) and Customs, habits, conventions (man made causes).

Features of Seasonal Variations

Following are the features of seasonal variations,

1. It indicates a variation which demonstrates periodical pattern of change in time series within a year.
2. It occurs for short repetitive calender period.
3. It may occur because of variations in temperature, rainfall, public holidays etc.

Q5. Components of Time Series

Answer : Jan.-21, Q4(OU)

The components of time series are as follows,

1. Secular Trend (or) Long Term Movements

Trend is the irreversible movement in a time series which continues generally in the same direction over a long period of time. In other words, secular trend refers to the general tendency of the time series data to increase or decrease over a long period of time. Trend refers to smooth, regular and long term movement of the data which has nothing to do with sudden and erratic movements either in upward or downward direction.

2. Seasonal Fluctuations

Seasonal fluctuations or variations refers to those periodic movements which occur regularly every year and have their origin in the nature of the year itself. They occur in regular and periodic manner over a span of less than a year, i.e., during a period of twelve months and have the same or almost the same pattern year after year. The causes of seasonal variations are climate in its widest sense (natural causes) and Customs, habits, conventions (man made causes).

3. Cyclical Fluctuations

The wave like movements in a time series with period of fluctuations for more than one year is called cyclical variations. They generally exhibit semi-regular periodicity as they are neither regular seasonal variations nor accidental erratic variations or fluctuations.

4. Random Erratic Fluctuations

Random erratic fluctuations refers to the fluctuations of long-term and short-term forces which is subjected to the occasional influences that may occur only once or several times without any pattern or regularity. A random variation may last a day or months.

Q6. Fit a trend line to the following data by the Freehand Method.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
Sales (₹)	19	22	24	20	23	25	23	26	25

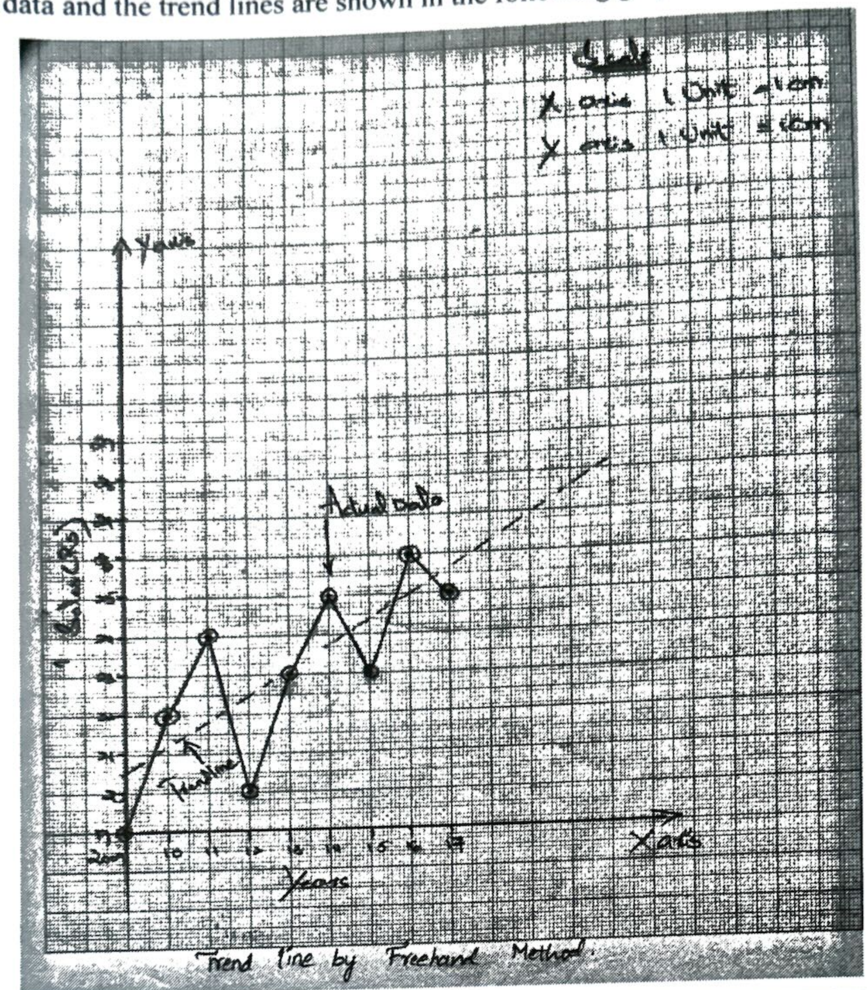
May/June-19, Q5 (OU)

Answer :

The given information is summarized below with graph.

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 lines are shown in the following graph.



Q7. 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

Jan.-21, Q5 (OU)

Answer :

Step-1

The trend mean values for the first 3 years are calculated as follows,

$$= \frac{20+16+24}{3} = \frac{60}{3} = 20$$

The trend mean values for the last 3 years are calculated as follows,

$$= \frac{30+28+32}{3} = \frac{90}{3} = 30$$

Therefore, the Semi-Averages are 20 and 30

Step-2

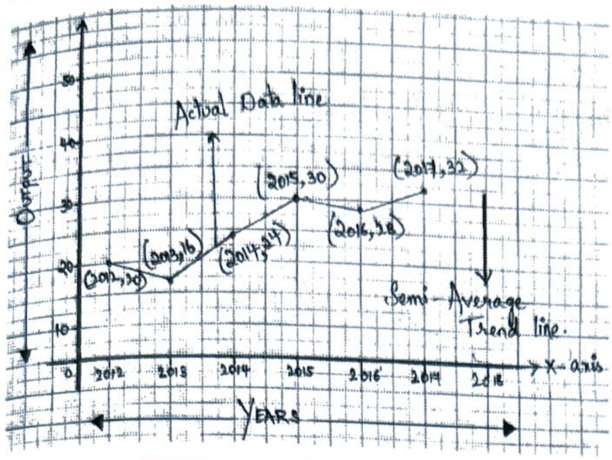
The next step is to plot the semi-averages against the mid-point (middle year) of each time period, thus, it would be year 2013 and 2016 respectively.

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,



Q8. Annual trend of Milk Consumption (Y) is 18.6 + 1.8X. Convert the equation into monthly basis.

Answer : March/April-17, Q7 (OU)

Given that,

$$Y = 18.6 + 1.8X$$

The annual trend equation $Y_c = a + bX$

The formula for converting annual trend into monthly trend is as follows,

$$Y_c = \frac{a}{12} + \frac{b}{12 \times 12} X$$

Substituting the values in the above equation, we get,

$$Y_c = \frac{18.6}{12} + \frac{1.8}{144} X$$

$$Y_c = 1.55 + 0.0125X$$

Q9. Given the following equation Y_c = 210 + 1.5X. Time origin is 2006. Time unit is one year, shift the origin to 2011.

Answer : Oct./Nov.-14, Q8 (OU)

Given that,

$$Y_c = 210 + 1.5X$$

Origin of year 2006 is to be shifted to year 2011. Time unit in 1 year, [2006 – 2011 = 5 years]

$$2006 = X$$

$$2007 = X + 1$$

$$2008 = X + 2$$

$$2009 = X + 3$$

$$2010 = X + 4$$

$$2011 = X + 5$$

$$Y_c = 210 + 1.5X$$

$$Y_c = a + b(X + K)$$

$$= 210 + 1.5(X + 5)$$

$$= 210 + 1.5X + 7.5$$

$$= 217.5 + 1.5X$$

Q10. Given the Trend Equation Y=34.5+2.1×[1999=0] change the base year to 2003 and re write the Trend Equation.

Answer : Sept./Oct.-21, Q7 (OU)

Given that,

$$Y = 34.5 + 2.1x$$

Origin of year 1999 to be change to the base year 2003 [2003 – 1999 = 4 years]

∴ Let the value of 1999 be x, so, the value of year 2003 will be $x + 4$.

Substitute $(x + 4)$ in the place of x

$$Y = 34.5 + 2.1(x + 4)$$

$$Y = 34.5 + 2.1x + 8.4$$

$$Y = 42.9 + 2.1x$$

∴ The trend equation for base year 2003 is $Y = 42.9 + 2.1x$

PART-B
ESSAY QUESTIONS AND ANSWERS

3.1 TIME SERIES – INTRODUCTION – COMPONENTS

Q11. What is Time Series? Explain the components of time series.

Answer :

Time Series

" A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables".

– *Ya-Lun Chou*

Time series refers to the arrangement of statistical data in chronological order (i.e.,) according to the time of occurrence. It represents the changing moments of variables over a particular period of time.

Time series plays an important role in business and economics. Thus, economists developed many statistical techniques for analyzing time series data. However, these techniques can also be applied to study time series of other disciplines which are not related to economics and statistics like natural sciences, social sciences, etc.

The functional relationship of time series can be mathematically represented as,

$$y = f(t)$$

Where, y = Variable under consideration

f = Functional relationship

t = Times $t_1, t_2, t_3, \ldots t_n$

Components of Time Series

The components of time series are as follows,

1. Secular Trend (or) Long Term Movements

Secular trend is the irreversible movement in a time series which continues generally in the same direction over a long period of time. In other words, secular trend refers to the general tendency of the time series data to increase or decrease over a long period of time. Trend refers to smooth, regular and long term movement of the data which has nothing to do with sudden and erratic movements either in upward or downward direction.

Example

(a) Upward trend : Price, incomes, population

(b) Downward trend: Deaths due to epidemics, bank interest rate, inflation rate.

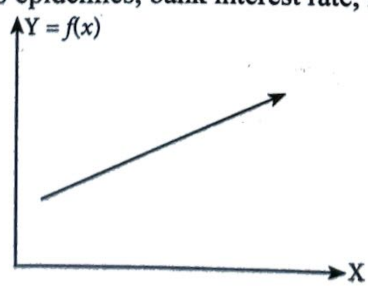


Figure: Secular Trend/Long Term Movements

2. Seasonal Fluctuations

Seasonal fluctuations or variations refers to those periodic movements which occur regularly every year and have their origin in the nature of the year itself. They occur in regular and periodic manner over a span of less than a year, i.e., during a period of twelve months and have the same or almost the same pattern year after year. The causes of seasonal variations are climate in its widest sense (natural causes) and customs, habits, conventions (man made causes).

Example

Time series influenced by seasonal variations are related to agricultural production, sales of agricultural produce, bank deposits, sales and profits in a departmental store.

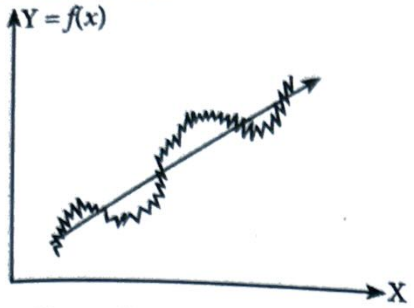


Figure: Seasonal Fluctuations

3. Cyclical Fluctuations

The wave like movements in a time series with period of fluctuations for more than one year are called cyclical variations. They generally exhibit semi-regular periodicity as they are neither regular seasonal variations nor accidental erratic fluctuations.

Example

Times of prosperity, production, sales, employment and other economic activities are high, and in times of depression it is low.

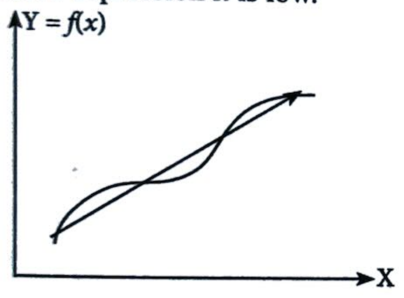


Figure: Cyclical Fluctuations

4. Random Erratic Fluctuations

Random erratic fluctuations refers to the fluctuations of long-term and short-term forces which is subjected to the occasional influences that may occur only once or several times without any pattern or regularity. A random variation may last a day or months.

Example

Wars, earthquakes, floods, fires, strikes, lockouts etc.

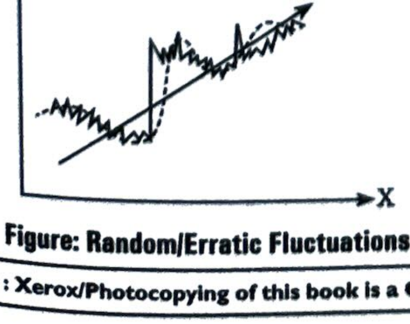


Figure: Random/Erratic Fluctuations

Q12. Write the features of tendency.

Answer : May/June-18, Q4(a) (KU)

The term trend means "Tendency". A trend refers to the general tendency of the time series data to increase or decrease over a long period of time It is also called as secular trend.

The following are the features of a trend/ tendency,

1. Long Period of Time

It is necessary to collect data for a long period of time in order to analyze a trend. Data points related to some period of time cannot be treated as a trend. But, the duration of the period mainly depends on the nature of the data. For example, in order to analyze the trend associated with economic development of a country, data such as population growth, agricultural production, software exports etc are required for at least 8 to 10 years. However, in case of readings taken in every 15 seconds over a period of 10 hours, may sufficient to study the effect of a medicine which regulates human pulse rate.

2. Upward, Downward or Stable Trend

Generally, the secular trend of a series is either upward or downward in nature. For example, data related to sales, production, income, population etc shows an upward trend. Similarly, data related to death rate, people below poverty line etc will shows a downward trend. However, for indices of stock market, a long term trend is upward, downward or stable over a period of time.

3. Impact of Stable Factors

The long term trend facilitates in capturing the impact of some forces which are less or more stable. Such factors fluctuate steadily and continuously over a period of time. For example, factors such as change in people's taste and customs, new material discovery, technology changes and so on. These factors do not reflect any sudden changes.

4. Linear or Non-Linear Trend

A trend can be either a linear or non linear when the time period is plotted on x-axis and value of variable on y-axis. A trend is said to be a linear if rate of increase or decrease is constant. Similarly, a trend can be non-linear or curvi-linear if the rate of growth is uneven or unpredictable. Thus, the plotted points on the graph do not result in a straight line.

3.2 METHODS OF MEASURING TREND

Q13. What are the various methods which can be used for measuring trend component of time series? Explain how trend is measured by using freehand curve with an example.

Answer :

The various methods which are used for measuring trend component of time series are,

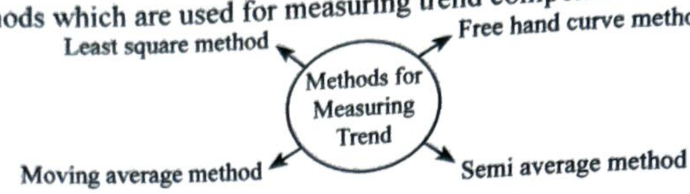


Figure: Methods for Measuring Trend

Free Hand Curve Method

Freehand curve refers to the trend which is determined by inspecting the plotted points on a graph sheet and observing the upward and downward movements of the points. It 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 change. Such a freehand smoothed curve eliminates the short-time swings and shows the long period general tendency of the changes in the data.

Example: Fit a trend line to the following data by the freehand method,

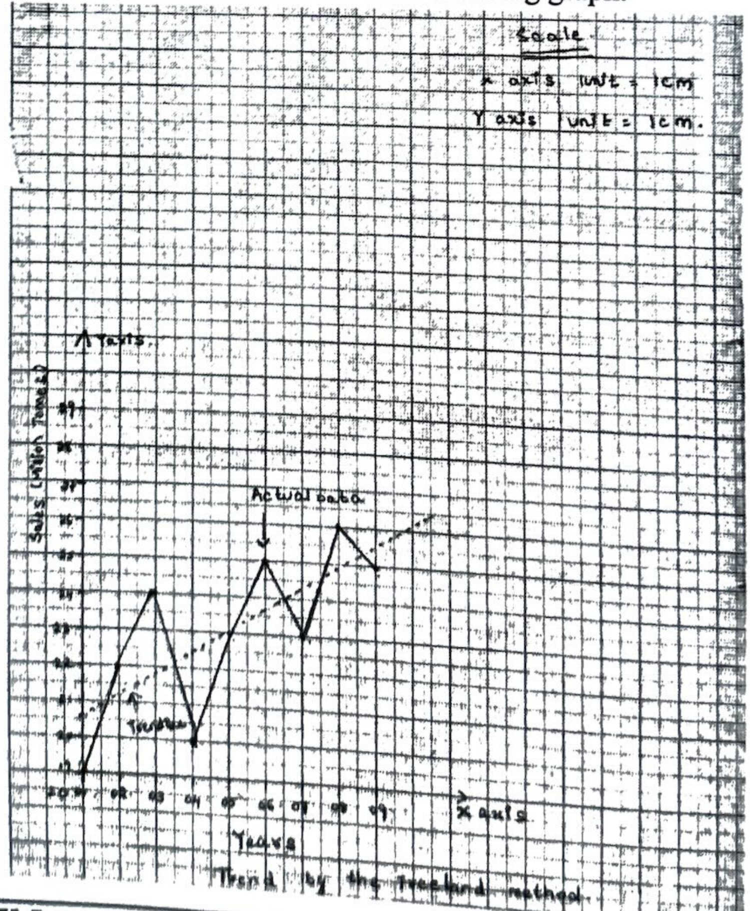
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sales (million tonnes)	19	22	24	20	23	25	23	26	25

Solution :

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 lines are shown in the following graph.



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. However different persons, draw 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.

3.2.1 Semi Averages Methods

Q14. Explain how trend analysis is done by semi averages method. State merits and demerits of semi average method.

Answer :

Trend Analysis through Semi Average Method

The average between two time periods is referred as semi average. The procedure followed for semi-average method is as follows,

Step-1: The entire time series is classified into two equal parts with respect to time. The even period is splitted into two equal parts while the odd periods are also classified into two equal parts but are obtained by omitting middle period.

Step-2: Compute the arithmetic mean of time series values for each half separately. These arithmetic means are called semi-averages.

Step-3: Semi averages are plotted as points against the middle point of the respective time periods covered by each part.

Step-4: 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,

1. It does not depend on personal judgement.
2. It is easy to apply and understand.
3. It helps to obtain past and future estimates by extending the line in both direction,.

Demerits of Semi-Average Method

The following are some of the demerits of semi-average method,

1. It assumes the presence of linear trend which may not exist.
2. It uses arithmetic mean for obtaining semi averages which is questionable.
3. Its 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

Solution :

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,

$$\text{Average of the first three years} = \frac{700+900+1100}{3} = \frac{2700}{3} = 900$$

$$\text{Average of the last three years} = \frac{1300+1000+1600}{3} = \frac{3900}{3} = 1300$$

Therefore, the semi-averages are $\boxed{900}$ and $\boxed{1300}$.

Step 2

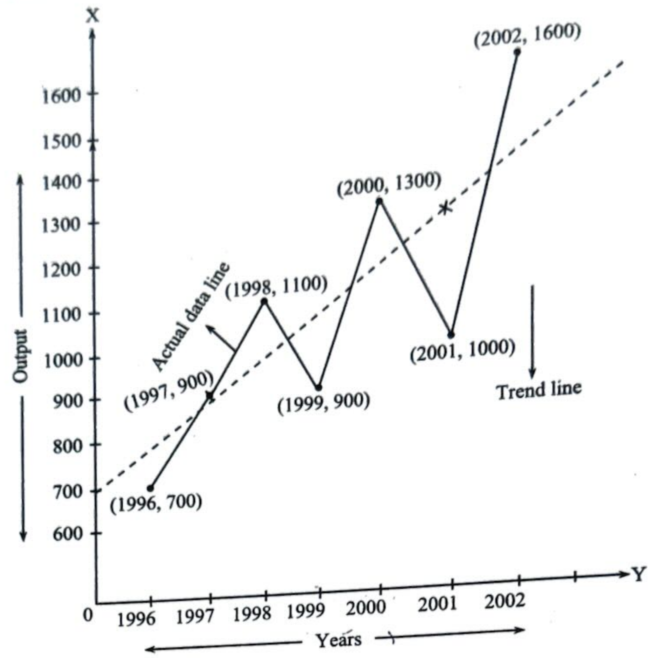
The next step is to plot the semi-averages against the mid-point (middle year) of each time period. Thus, it would be year 1997 and 2001 respectively.

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,



3.2.2 Moving Averages Method

Q15. What is Moving Average Method? Discuss the method of moving averages in measuring trend. What are its merits and limitations?

Answer :

Moving Average Method

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 length of the cycle. It is more applicable to data with cyclical movements.

Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3} \ldots$$

Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5} \ldots$$

Methods of Moving Average

The following are the two methods of moving averages,

(i) Odd Yearly Method

Step-1: Calculate 3/5...yearly totals

Step-2: Now compute 3/5 yearly average by dividing the totals calculated in step-1 by the respective number of years. i.e., 3/5/...

Step-3: Short term oscillations are calculated using the formula,

$$Y - Y_c$$

(ii) Even Yearly Method

Example: 4 years

Step-1: Calculate 4 yearly moving totals and place at the centre of middle two years of the four years considered.

Step-2: Divide 4 yearly moving totals by 4 to get 4 yearly average.

Step-3: Take a 2 period moving average of the moving average which gives the 4 yearly moving average centered.

Merits of Moving Average

The merits of moving average are as follows,

1. It is a simplest method among all mathematical methods of fitting trend.
2. It is a flexible method, even if a few more observations are to be added, the entire calculations are not changed.
3. It coincides with the period of the cycle. Thus, cyclical fluctuations are automatically eliminated.
4. Its curve is determined by the data rather than the statisticians choice of mathematical function.

Limitations of Moving Average

The following are the limitations of moving averages,

1. Its 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.
2. It is difficult to decide the period of moving average since there is no hard and fast rule for the purpose.
3. It cannot be used in forecasting as it is not represented by any mathematical function.
4. It lies either above or below the true sweep of the data when the trend is not linear.

PROBLEMS ON MOVING AVERAGE METHOD

Q16. Calculate a 7-year Moving Average for the following data on the number of commercial and industrial failures in a country during 1987-2002.

Year	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
No. of failures	23	26	28	32	20	12	12	10	9	13	11	14	12	9	3	1

Solution :

Sept./Oct.-21, Q13 (OU)

Calculation of Trend by 7–yearly Moving Average

Year (1)	No. of Failures (2)	7 Yearly Moving Totals (3)	7 Year Moving Average (4)
1987	23	–	–
1988	26	–	–
1989	28	–	–
1990	32	153	21.86
1991	20	140	20
1992	12	123	17.57
1993	12	108	15.43
1994	10	87	12.43
1995	9	81	11.57
1996	13	81	11.57
1997	11	78	11.14
1998	14	71	10.14
1999	12	63	9
2000	9	–	–
2001	3	–	–
2002	1	–	–

Q17. 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

Jan.-21, Q13 (OU)

Solution :

Calculation of 3-yearly and 5-yeraly Moving Averages

Years	Production (In Quintals)	3-Yearly Moving Total	3-Yearly Moving Averages	5-Yearly Moving Total	5-Yearly Moving Averages
2006	500	-	-	-	-
2007	540	500 + 540 + 550 = 1,590	$\frac{1590}{3} = 530$	-	-
2008	550	540 + 550 + 530 = 1620	$\frac{1620}{3} = 540$	500 + 540 + 550 + 530 + 520 = 2640	$\frac{2640}{5} = 528$
2009	530	550 + 530 + 520 = 1600	$\frac{1600}{3} = 533.33$	540 + 550 + 530 + 520 + 560 + 2700	$\frac{2700}{5} = 540$
2010	520	530 + 520 + 560 = 1610	$\frac{1610}{3} = 536.66$	550 + 530 + 520 + 560 + 600 = 2760	$\frac{2760}{5} = 552$
2011	560	520 + 560 + 600 = 1680	$\frac{1680}{3} = 560$	530 + 520 + 560 + 600 + 640 = 2850	$\frac{2850}{5} = 570$
2012	600	560 + 600 + 640 = 1800	$\frac{1800}{3} = 600$	520 + 560 + 600 + 640 + 620 = 2940	$\frac{2940}{5} = 588$
2013	640	600 + 640 + 620 = 1860	$\frac{1860}{3} = 620$	560 + 600 + 640 + 620 + 610 = 3030	$\frac{3030}{5} = 606$
2014	620	640 + 620 + 610 = 1870	$\frac{1870}{3} = 623.33$	600 + 640 + 620 + 610 + 640 = 3110	$\frac{3110}{5} = 622$
2015	610	620 + 610 + 640 = 1870	$\frac{1870}{3} = 623.33$	-	-
2016	640	-	-	-	-

Q18. Calculate three year moving average for the following data:

Year:	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Value	242	250	252	249	253	255	251	257	260	265	262

July/Aug.-21, Q1(c) (KU)

Solution :

Calculation of 3 years Moving Averages

Year	Value	3-years Moving total	3-years Moving Average
2008	242	-	-
2009	250	242+250+252=744	$\frac{744}{3} = 248.00$
2010	252	250+252+249=751	$\frac{751}{3} = 250.33$
2011	249	252+249+253=754	$\frac{754}{3} = 251.33$
2012	253	249+253+255=757	$\frac{757}{3} = 252.33$
2013	255	253+255+251=759	$\frac{759}{3} = 253.00$
2014	251	255+251+257=763	$\frac{763}{3} = 254.33$
2015	257	251+257+260=768	$\frac{768}{3} = 256.00$
2016	260	257+260+265=782	$\frac{782}{3} = 260.67$
2017	265	260+265+262=787	$\frac{787}{3} = 262.33$
2018	262	-	-

Q19. 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

(May/June-19, Q11(a) (OU) | Oct./Nov.-12, Q13(a) (OU))

Solution :

Calculation of Trend by 4-yearly Moving Averages

Year (1)	Production (In Tonnes) (2)	4 yearly Moving Totals (3)	4 yearly Moving Averages (4) = (3) ÷ 4	2 Point Moving Totals (5)	4 yearly Moving Average Centered (6) = (5) ÷ 2
2009	506	–			
2010	620	–			
2011	1036	2835	708.75		
2012	673	2917	729.25	1438	719
2013	588	2993	748.25	1477.5	738.75
2014	696	3073	768.25	1516.50	758.25
2015	1116	3138	784.5	1552.75	776.375
2016	738	3213	803.25	1587.75	793.875
2017	663	–			

3.2.3 Least Square Method

Q20. What is the least square method? What are the merits and demerits of this method?

OR

What is the least square method and explain its advantages and disadvantages?

Answer : May/June-19, Q8(b) (MGU)

Least Square Method

The least square method is a statistical procedure which is used to find the best fit curve for the set of data where different variables are involved. This method is mostly used for the time series of data in which the relationship of two or more variables is difficult to identify. Least square method provides a trend line of best fit in the form of curve in order to represents the relationship between a known and unknown variable. Because of this reason the trend line is also called as line of best fit. This line can be a straight line trend or parabolic trend through which sum of squares of the distance from different points is reduced or minimized.

The straight line trend equation is in the form of $Y_c = a + bX$

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

$$a = \frac{\Sigma Y}{N} \; ; b = \frac{\Sigma XY}{\Sigma X^2}$$

By substituting a and b values in straight line trend equation $Y_c = a + bX$, we get the straight line equation which can be used for estimation of future values.

Merits/Advantages of Least Squares

The following are the merits of least squares method,

1. It is a mathematical method of measuring trend and is free from subjectiveness.
2. It 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. It enables us to compute the trend values for all the given time periods in the series.
4. It is the equation which can be used to estimate the values of the variable for any given time period 't' in future and the forecasted values are quite reliable.
5. It is the only technique which enables us to obtain the rate of growth per annum for yearly data in case of linear trend.

Demerits/Disadvantages of Least Squares

The demerits of least squares are as follows,

1. It becomes necessary to do fresh calculation even if a single new observation is added.
2. Its calculations are quite tedious and time consuming as compared with other methods.
3. Its future predictions completely ignore the cyclical, seasonal and erratic fluctuations.
4. It 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 ON LEAST SQUARES METHOD (TREND LINE)

Q21. Use the Method of Least Squares to fit a Straight Line Trend to the following data.

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Avg. Monthly Demand	61	66	72	76	82	90	96	100	103	110	114

Solution : Sept./Oct.-21, Q14 (OU)

The straight line trend equation is $y_c = a + bx$ by solving this normal equations, we get a and b.

Normal equations are,

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

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

Fitting the Straight Line Trend

Year	Average Monthly Demand (y)	Deviations from 1997 (x)	x^2	xy
1992	61	–5	25	–305
1993	66	–4	16	–264
1994	72	–3	9	–216
1995	76	–2	4	–152
1996	82	–1	1	–82
1997	90	0	0	0
1998	96	1	1	96
1999	100	2	4	200
2000	103	3	9	309
2001	110	4	16	440
2002	114	5	25	570
N=11	$\Sigma y = 970$	$\Sigma x = 0$	$\Sigma x^2 = 110$	$\Sigma xy = 596$

Since, $\Sigma x = 0$

$$a = \frac{\Sigma y}{N} = \frac{970}{11} = 88.18$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{596}{110}\ 5.418$$

∴ The Straight line trend $y_e = a + b\,x$ is,

$y_e = 88.18 + 5.418\,x$

Q22. 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

Solution : Jan.-21, Q14 (OU)

Equation for straight line trend is $Y_c = a + bx$. The value of a and b can be attained by solving the following two normal equations,

$\Sigma y = Na + b\Sigma x$ ....(1)

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

Now, fitting of straight line trend by the method of least squares.

Year	Sales (in '000 units) y	2013 (x)	x^2	xy	Trend Value Y_c (WN)
2010	140	– 3	9	– 420	139.42
2011	144	– 2	4	– 288	146.28
2012	160	– 1	1	– 160	153.14
2013	152	0	0	0	160
2014	168	1	1	168	166.86
2015	176	2	4	352	173.72
2016	180	3	9	540	180.58
N = 7	$\Sigma y = 1120$	$\Sigma x = 0$	$\Sigma y^2 = 28$	$\Sigma xy = 192$	$\Sigma y_c = 1120$

Working Notes (WN)

$\Sigma y_e = Na + b\Sigma x$

$1120 = 7a + 6(0)$

Since the value of $\Sigma x = 0$, $\Sigma y = Na$

Based on the above table calculation, calculating value of a and b

$$a = \frac{\Sigma y}{N} = \frac{1120}{7} = 160$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{192}{28} = 6.86$$

Hence, the value of $a = 160$, $b = 6.86$

The equation of straight line trend is,

$Y_e = 160 + 6.86\,x$

2010 ⇒ When $x = -3$, $y = 160 + 6.86\,(-3)$

$= 160 - 20.58$

$= 139.42$

2011 ⇒ When $x = -2$, $y = 160 + 6.86\,(-2)$

$= 160 - 13.72$

$= 146.28$

2012 ⇒ When $x = -1$, $y = 160 + 6.86\ (-1)$
$= 160 - 6.86$
$= 153.14$

2013 ⇒ When $x = 0$, $y = 160 + 6.86\ (0)$
$= 160 - 0$
$= 160$

2014 ⇒ When $x = 1$, $y = 160 + 6.86\ (1)$
$= 160 + 6.86$
$= 166.86$

2015 ⇒ When $x = 2$, $y = 160 + 6.86\ (2)$
$= 160 + 13.72$
$= 173.72$

2016 ⇒ When $x = 3$, $y = 160 + 6.86\ (3)$
$= 160 + 20.58$
$= 180.58.$

Q23. Below are given the figures of production (in thousand quintals) of a sugar factory.

Years	2001	2002	2003	2004	2005	2006	2007
Production	77	88	94	85	91	98	90

(i) Fit a straight line by the least squares method and tabulate the trend values.

(ii) What is the yearly increase in the production of sugar? March/April-14, Q13(b) (OU)

OR

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

May/June-19, Q11(b) (OU)

Solution :

(i) Equation for straight line trend is $Y_c = a + bX$.

The value of a and b can attained by solving the following two normal equations,

$$\Sigma Y_c = Na + b\Sigma X \quad \text{... (1)}$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2 \quad \text{... (2)}$$

Fitting of straight line trend by the method of least squares.

Year	Production Y	2004 (X) (Origin)	X^2	XY	Trend Value Y_c
2001	77	–3	9	–231	83
2002	88	–2	4	–176	85
2003	94	–1	1	–94	87
2004	85	0	0	0	89
2005	91	1	1	91	91
2006	98	2	4	196	93
2007	90	3	9	270	95
N = 7	$\Sigma Y = 623$	$\Sigma X = 0$	$\Sigma X^2 = 28$	$\Sigma XY = 56$	$\Sigma Y_C = 623$

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

$623 = 7a + b(0)$

Since value of $\Sigma X = 0$, $\Sigma Y = Na$

$$a = \frac{\Sigma Y}{N} = \frac{623}{7} = 89$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{56}{28} = 2$$

Hence, the value of $a = 89$ and $b = 2$

The equation of straight line trend is, $Y_e = 89 + 2X$

When $X = -3$, $Y = 89 + 2(-3) = 89 - 6 = 83$

When $X = -2$, $Y = 89 + 2(-2) = 89 - 4 = 85$

When $X = -1$, $Y = 89 + 2(-1) = 89 - 2 = 87$

When $X = 0$, $Y = 89$

When $X = 1$, $Y = 89 + 2(1) = 89 + 2 = 91$

When $X = 2$, $Y = 89 + 2(2) = 89 + 4 = 93$

When $X = 3$, $Y = 89 + 2(3) = 89 + 6 = 95$

(ii) The yearly increase in the production of sugar is $b = 2$ (in thousand quintals).

Q24. Fit a trend equation y = a + bx

Year	2007	2008	2009	2010	2011	2012	2013	2014
Production	50	67	76	85	92	87	90	112

Answer : July/Aug.-21, Q5 (MGU)

Since the observations are even (N = 8). Shift the origin to the middle of the two periods i.e., 2010 and 2011. Therefore, the values of r can be calculated as,

$$r = \frac{t - \left(\frac{2010 + 2011}{2}\right)}{\frac{1}{2}(interval)} \Rightarrow 2(t - 2010.5)$$

$\therefore r = t - 2010.5$

Fitting of trend values by the method of least squares

Year (t)	Production (000 tones) (Y)	x = 2 (t – 2010.5)	xy	x^2	Trend values $Y_C = 82.3 + (-3.50)\,x$
2007	50	–7	–350	49	106.87
2008	67	–5	–335	25	99.87
2009	76	–3	–228	9	92.87
2010	85	–1	–85	1	85.87
2011	92	1	92	1	78.87
2012	87	3	261	9	71.87
2013	90	5	450	25	64.87
2014	112	7	784	49	57.87
N = 8	$\Sigma y = 659$	$\Sigma x = 0$	$\Sigma xy = 589$	$\Sigma x^2 = 168$	$Y_c = 658.96$

$Y_c = a + bx$

$\therefore a = \frac{\Sigma y}{n} = \frac{659}{8} = 82.37$

$\therefore b = \frac{\Sigma xy}{x^2} = \frac{589}{168} = 3.50$

The straight line trend is,

$Y_c = 82.37 + (-3.50)$

When x = – 7, y = 82.37 – 3.50 (– 7) = 82.37 + 24.5 = 106.87

When x = – 5, y = 82.37 – 3.50 (– 5) = 82.37 + 17.5 = 99.87

When x = – 3, y = 82.37 – 3.50 (– 3) = 82.37 + 10.5 = 92.87

When x = – 1, y = 82.37 – 3.50 (– 1) = 82.37 + 3.5 = 85.87

When x = 1, y = 82.37 – 3.50 (1) = 82.37 – 3.5 = 78.87

When x = 3, y = 82.37 – 3.50 (3) = 82.37 – 10.5 = 71.87

When x = 5, y = 82.37 – 3.50 (5) = 82.37 – 17.5 = 64.87

When x = 7, y = 82.37 – 3.50 (7) = 82.37 – 24.5 = 57.87

Q25. You are given below the figures of production of a fertilizer factory (in thousand tonnes).

Year	2005	2006	2007	2008	2009	2010	2011	2012
Production	140	150	180	196	170	182	200	220

(i) Fit a straight line trend using the method of least squares and tabulate the trend values.

(ii) What is the monthly increase in population?

(iii) Estimate the production for the year 2015.

Answer :

July/Aug.-21, Q6 (MGU)

Since the observations are even (N = 8). Shift the origin to the middle of the two line periods i.e., 2008 and 2009. Therefore, the values of x can be calculated as,

$$x = \frac{t - \left(\frac{2008 + 2009}{2}\right)}{\frac{1}{2}(interval)} \Rightarrow 2(t - 2008.5)$$

$\therefore x = t - 2008.5$

Fitting of Straight Line Trend by the Method of Least Squares

Year (t)	Production (Y)	X = 2 (t – 2008.5)	xy	x^2	Trend values Y_c
2005	140	– 7	– 980	49	212.65
2006	150	– 5	– 750	25	203.25
2007	180	– 3	– 540	9	193.85
2008	196	– 1	– 196	1	226.75
2009	170	1	170	1	132.75

2010	182	3	546	9	165.65
2011	200	5	1,000	25	156.25
2012	220	7	1,540	49	146.85
N = 8	$\Sigma y = 1,438$	$\Sigma x = 0$	$\Sigma xy = 790$	$\Sigma x^2 = 168$	$Y_c = 1,438$

$$\therefore a = \frac{\Sigma y}{n} = \frac{1,438}{8} = 179.75$$

$$\therefore b = \frac{\Sigma xy}{x^2} = \frac{790}{168} = 4.70$$

$\therefore$ The straight line trend is,

$y_e = a + bx$

$y_e = 179.75 + (-4.70)\, x$

When x = − 7, y = 179.75 − 4.70 (− 7) = 179.75 + 32.9 = 212.6

When x = − 5, y = 179.75 − 4.70 (− 5) = 179.75 + 23.5 = 203.25

When x = − 3, y = 179.75 − 4.70 (− 3) = 179.75 + 14.1 = 193.85

When x = − 1, y = 179.75 − 4.70 (− 1) = 179.75 + 47 = 226.75

When x = 1, y = 179.75 − 4.70 (1) = 179.75 + 47 = 132.75

When x = 3, y = 179.75 − 4.70 (3) = 179.75 + 14.1 = 165.65

When x = 5, y = 179.75 − 4.70 (5) = 179.75 + 23.5 = 156.25

When x = 7, y = 179.75 − 4.70 (7) = 179.75 + 32.9 = 146.85

(ii) Monthly Increase in production

$4.70 \times 1000 = 4,700$

Hence, the monthly increase in production is,

$$= \frac{4,700}{12}$$

= 391.666 tons

(iii) Estimate the production for the year 2015.

Deviation x for the year 2015 can be calculated as,

$x = 2\,(t - 2008.5)$

$= 2\,(2015 - 2008.5)$

$= 4030 - 4017$

$= 13,$

$y = 179.75 - 4.70\,(13)$

$= 179.75 - 61.6$

$= 118.15$

Thus, the production for the year 2015 is 118.15.

3.3 DESEASONALISATION OF DATA

Q26. What do you mean by deseasonalisation of data? Explain it with an example problem.

Answer :

Deseasonalisation of Data

Deseasonalisation of data refers to the process of eliminating seasonal fluctuations from the given time series. It usually facilitates in adjusting the given time series for seasonal fluctuations and therefore remains with variables like trend component, cyclical fluctuations and irregular variations.

In multiplicative model of time series, deseasonalised values are calculated by dividing the given values with the respective indices of seasonal variations.

$$\text{Deseasonalised data} = \frac{Y}{S} = \frac{TCSI}{S} = TCI$$

In additive model of time series, deseasonalised values are calculated by deducting seasonal variations from the given values,

$$\text{Deseasonalised data} = Y - S = (T+S+C+I) - S = T+C+I$$

Example

'XYZ' Co. estimates its average sales in a particular year to be ₹ 2,00,000. The seasonal indices of the sales data are as follows.

Month	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Seasonal	95	60	100	98	106	110	120	136	125	96	75	79

Using this information calculate monthly sales for the company. Assuming that there is no trend.

Solution :

$$\text{Seasonal effect} = \frac{\text{Seasonal index}}{100}$$

Average annual sale = ₹ 2,00,000

∴ Average monthly sales = 2,00,000 × Seasonal effect

Month (1)	Seasonal Indices (2)	Seasonal Effect $(3) = \frac{(2)}{100}$	Monthly Sales (4) = (3) × 2,00,000
January	95	[95/100] 0.95	1,90,000
February	60	[60/100] = 0.6	1,20,000
March	100	[100/100] = 1	2,00,000
April	98	[98/100] = 0.98	1,96,000
May	106	[106/100] = 1.06	2,12,000
June	110	[110/100] = 1.1	2,20,000
July	120	[120/100] = 1.2	2,40,000
August	136	[136/100] = 1.36	2,72,000

September	125	[125/100] = 1.25	2,50000
October	96	[96/100] = 0.96	1,92,000
November	75	[75/100] = 0.75	1,50,000
December	79	[79/100] = 0.79	1,58,000
	1200	**12**	**24,00,000**

3.4 USES AND LIMITATIONS OF TIME SERIES

Q27. Discuss the uses and limitations of Time series.

Answer : July/Aug-21, Q6(KU)

Uses/Utility of Time Series

The analysis of time series is not only used by the economists and business men but it is also followed and used by the scientists, astronomists, geologists, sociologists, biologists and researchers. The uses or utility of time series analysis are as follows,

1. **Understanding Past Behaviour**

   It helps in understanding the past behaviour by considering the changes that have taken place in the past. They predict the future behaviour with the help of past data.

2. **Planning Future Operations**

   It helps in future planning by forecasting the events and their relationship. If regular occurrence of any event is there over a long period then such event is considered to predict the future.

3. **Evaluating Current Accomplishments**

   It helps in evaluating the performance by comparing the actual performance with the expected performance and analyse the reasons of variations if any. For example, If expected sale of a refrigerator is 10,000 for 2011-12 but actual sale is only 9000. By using time series analysis, they can evaluate the reason for it's shortfall.

4. **Facilitates Comparison**

   It compare the different periods and various conclusions are drawn upon it. But it is not necessary that everyone should believe on it. As statisticians are not foretellers, they cannot predict 100% accurate results for future events.

   The future prediction could be possible only if they include the influences of various forces which will effect the series such as climate, customs and traditions and other factors like growth and declining factors and complex forces that effects the production of business cycles and such analysis is examined carefully for number of times.

Limitations of Time Series

The limitations of time series are as follows,

1. It may not be used/applied in all the situations due to insufficient data.
2. It takes into consideration the environmental factors which result in variations.
3. It may result into errors or deviations, which needs to be monitored.
4. It is difficult to predict cyclical and random variations.
5. When cyclical and random variations are ignored, forecasts based on the extension of the trend line and seasonal indices may lead to inaccuracy in some cases.

EXERCISE AND PRACTICE QUESTIONS

SHORT QUESTIONS

THEORY

Q1. What is Time Series? [Refer, Q1] May/June-19, Q4 (OU)

Q2. Utility of Time Series Analysis. [Refer, Q2]

OR May/June-18, Q4 (OU)

What are the uses of time series? May/June-19, Q3 (MGU)

Q3. Objectives of Time Series [Refer, Q3]

Q4. What is seasonal variations? Write three features of seasonal variations. [Refer, Q4] May/June-18, Q1(c) (KU)

Q5. What are the Components of Time Series? [Refer, Q5]

PROBLEMS

Q6. Fit a trend line to the following data by the Freehand Method. [Refer Similar, Q6]

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sales (₹)	20	23	25	21	24	26	24	27	26

Q7. Annual trend of Milk Consumption (Y) is 20.6 + 1.9 X. Convert the equation into monthly basis. [Refer Similar, Q8]

(Ans: $Y_c = 1.717 + 0.0132X$)

Q8. Calculate 3 yearly moving averages from the following data, [Refer Similar, Q7]

Year	2013	2014	2015	2016	2017	2018	2019
Sales (in million ₹)	82	88	90	92	94	96	98

(Ans: 2014 = 86.66 ; 2015 = 90 ; 2016 = 92 ; 2017 = 94 ; 2018 = 96)

Q9. Given the following equation $Y_e = 420 + 3X$. Time origin is 2006. Time unit is one year, shift the origin to 2011. [Refer Similar, Q8]

(Ans: $Y_c = 427.5 + 3X$)

Q10. Calculate five yearly moving averages for the following data: [Refer Similar, Q7]

Year	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
No. of students	332	317	357	392	402	405	410	417	405	431

March/April-16, Q8(OU)

(Ans: 5-yearly M.A's for 1983 to 1988 respectively are: 360, 374.6, 393.2, 405.2, 407.8, 413.6).

Q11. Given Y = 300 + 24 X (origin 2001) X-unit = 1 year. Y-unit = Annual profits (₹1000). Convert into monthly trend equation. [Refer Similar, Q8] March/April-11, Q8 (OU)

(Ans: $Y_c = 25 + 0.1667X$)

Q12. Given Y = 150 + 12X (origin 2009) X-unit = 1 year, Y-unit = annual profits ('000 ₹). Convert into monthly trend equation. [Refer Similar, Q8] Sept./Oct.-15, Q8 (OU)

(Ans: $Y_c = 12.5 + 0.08333X$)

ESSAY QUESTIONS

THEORY

Q1. What is Time Series? Explain the components of time series. [Refer, Q11]

Q2. Write the features of tendency. [Refer, Q12] May/June-18, Q4(a) (KU)

Q3. State the uses and limitations of time series. [Refer, Q27]

PROBLEMS

TREND EQUATION OR TREND LINE

Q4. Compute the trend values by the method of semi-averages from the data given below, [Refer Similar, Q18]

Year	1992	1993	1994	1995	1996	1997	1998	1999
No. of sheep (in lakhs)	56	55	51	47	42	38	35	32

(Ans: Trend values (in lakhs) for the years 1992 to 1999 are: 59, 56, 50.5, 46.5, 41.5, 37, 32.5, 28).

Q5. The sale of a commodity in tonnes varied from January 1999 to December 1999 in the following manner: [Refer Similar, Q23]

280	300	280	280	270	240
230	230	220	200	210	200

Find a trend by the method of semi-average.

Fit a trend line from the following data by using semi-average method:

Year	1993	1994	1995	1996	1997	1998
Profits (in '000)	100	120	140	150	130	200

(Ans: Joining the points (1994, 120) and (1997, 160), we get the trend line).

Q6. Fit a straight line trends to the following data and estimate the likely profit for the year 2004. [Refer Similar, Q25] March/April-16, Q13(b) (OU)

Year	1995	1996	1997	1998	1999	2000	2001
Profits in lakhs rupee	60	72	75	65	80	85	95

(Ans: $Y_c = 76 + 4.857X$ Likely profit for the year 2004 is ₹ 105.142 lakhs)

Q7. Fit straight line trend to the following data by using least square method. [Refer Similar, Q25]

Year	2006	2007	2008	2009	2010	2011	2012
Production (in lakhs of tons)	48	50	58	52	45	41	49

(Ans: $Y_c = 49 + (-1)X$) Sept./Oct.-15, Q13(a) (OU)

Q8. Fit a straight line trend by the method least squares to the following data and also predict the earnings for the year 2013, [Refer Similar, Q22] Oct./Nov.-14, Q13(a) (OU)

Year	2005	2006	2007	2008	2009	2010	2011	2012
Earnings (₹ in lakhs)	38	40	65	72	69	60	87	95

(Ans: $Y_e = 65.75 + 3.66\ X$, Predicted earnings for the year 2013 is 98.69 lakhs).

Q9. Production figure of a Textile Industry are as follows, [Refer Similar, Q25] May/June-18, Q11(b) (OU)

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.

(Ans: $Y_c = 13 + 0.75X$)

Q10. Find the straight line tendency for the following data using least square method.

Year	2010	2011	2012	2013	2014	2015
Production	62	83	90	80	90	95

(Ans: $Y_c = 83.33 + 2.75X$) [Refer Similar, Q25] May/June-18, Q4(b) (KU)

Q11. From the following data calculate trend values based on least square method and estimate production in 1993. [Refer Similar, Q21] May/June-19, Q8(a) (MGU)

Year	1982	1983	1984	1985	1986	1987	1988	1989
Production (000 tonnes)	58	56	55	51	47	38	35	32

(Ans: $Y_c = 46.5 + (-2.04)X$)

Q12. Fit a straight line trend by least squares method and estimate trend for 2015.

Year	2009	2010	2011	2012	2013	2014
Sales (₹ in ' 000s)	10	12	15	16	18	19

(Ans: $Y_c = 15 + 0.91X$) [Refer Similar, Q22] March/April-17, Q13(a) (OU)

Q13. The following data relate to the number of passenger cars (in millions) sold from 2004 to 2011. [Refer Similar, Q22] March/April-13, Q13(b) (OU)

Years	2004	2005	2006	2007	2008	2009	2010	2011
Number	6.7	5.3	4.3	6.1	5.6	7.9	5.8	6.1

(a) Fit a straight line trend to the data through 2009.

(b) Use your result in, (a) To estimate production in 2011 and compare with the actual production.

(Ans: $Y_c = 6.128 + 0.102X$)

MOVING AND SEMI-AVERAGE

Q14. Calculate 3 yearly moving averages from the following data, [Refer Similar, Q24]

Year	2007	2008	2009	2010	2011	2012	2013
Sales (In million ₹)	41	44	45	46	47	48	49

(Ans: 43.33, 45, 46, 47, 48). March/April-15, Q7 (OU)

Q15. Calculate 5 yearly moving averages from the following data. [Refer Similar, Q17]

Year	2002	2003	2004	2005	2006	2007
Production	100	105	115	90	95	85
Year	2008	2009	2010	2011	2012	2013
Production	80	65	75	70	75	80

March/April-15, Q13(a) (OU)

(Ans: 5-yearly moving averages are 101, 98, 93, 83, 80, 75, 73, 73).

Q16. Draw a trend line by the method of semi-averages from the following data,

Year	1984	1985	1986	1987	1988	1989	1990	1991
Sales '000 (unit)	200	120	128	192	204	126	224	228

Also predict the sales from the year 1993 from the graph. [Refer Similar, Q18]

(Ans: Sales for 1993 = 230).

Q17. Find the 4 yearly moving averages from the following data: [Refer Similar, Q18]

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Production (in Tonnes)	150	170	196	180	190	216	248	280	300	320

(Ans: 4 yearly moving averages = 274). May/June-18, Q11(a) (OU)

INTERNAL ASSESSMENT/EXAM

I Multiple Choice

1. _________ helps in evaluating current accomplishments. [ ]
   (a) Correlation (b) Time series
   (c) Trend analysis (d) Index numbers
2. _________ refers to only smooth, regular, long-term movement of the data. [ ]
   (a) Trend (b) Random variation
   (c) Cyclical variation (d) None of the above
3. Deaths due to epidemics, bank interest rate, inflation rate etc are the examples of _________. [ ]
   (a) Upward trend (b) Downward trend
   (c) Both (a) and (b) (d) All the above
4. Which of the following are the types of periodic movements? [ ]
   (a) Seasonal fluctuations (b) Cyclical fluctuations
   (c) Both (a) and (b) (d) Secular fluctuations
5. Methods which can be used for measuring trend component of time series are _________. [ ]
   (a) Semi averages (b) Moving averages
   (c) Least square (d) All the above
6. Short-term oscillations are calculated by using the _________ formula. [ ]
   (a) $Y + Y_e$ (b) $Y \div Y_e$
   (c) $Y - Y_e$ (d) $Y \times Y_e$
7. The straight line trend equation is in the form of _________ [ ]
   (a) $X = a + bX$ (b) $Y = a + bX$
   (c) $X = a - bX$ (d) $Y = a - bX$
8. _________ method is used to fit either a straight line trend or a parabolic trend. [ ]
   (a) Least square (b) Semi averages
   (c) Freehand curve (d) Moving average
9. Seasonal Effect = _________ [ ]
   (a) $\frac{\text{Seasonal Index}}{50}$ (b) $\frac{\text{Seasonal Index}}{10}$
   (c) $\frac{\text{Seasonal Index}}{30}$ (d) $\frac{\text{Seasonal Index}}{100}$
10. In multiplicative model of time series deseasonalised values are calculated by _________ the given values with the respective indices of seasonal variations. [ ]
   (a) Dividing (b) Multiplying
   (c) Adding (d) Deducting

II Fill in the Blanks

1. __________ is an arrangement of statistical data in a chronological order i.e., in accordance with its time of occurrence.
2. __________ refers to the general tendency of the time series data to increase or decrease over a long period of time.
3. The wave like movements in a time series with period of oscillation more than one year are called __________.
4. Wars, earthquakes, floods, fires, strikes, lockouts etc., are the examples of __________.
5. Under __________ method, the average value for a number of years is secured and this average is taken as the normal or trend value.
6. __________ 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.
7. $\Sigma XY =$ __________.
8. __________ refers to the process of eliminating seasonal fluctuations from the given time series.
9. In addictive model of time series, deseasonalised values = __________.
10. __________ are plotted as points against the middle point of the respective time periods covered by each part.

KEY

I. Multiple Choice

1. (b)
2. (a)
3. (b)
4. (c)
5. (d)
6. (c)
7. (b)
8. (a)
9. (d)
10. (a)

II. Fill in the Blanks

1. Time series
2. Secular trend
3. Cyclical fluctuations
4. Random fluctuations
5. Moving average
6. Least squares
7. $a\Sigma X + b\Sigma X^2$
8. Deseasonalisation of data
9. $Y - S = (T + S + C + I) - S = T + C + I$
10. Semi averages.

III Very Short Questions and Answers

Q1. Define Time Series.

Answer :

" A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables".

– Ya-Lun Chou

Time series refers to the arrangement of statistical data in chronological order i.e., according to the time of occurrence. It represents the changing moments of variables over a particular period of time.

Q2. What is free hand curve method?

Answer :

Freehand curve refers to the trend which is determined by inspecting the plotted points on a graph sheet and observing the upward and downward movements of the points. It smooth out the irregularities by drawing a freehand curve or line through the scatter points.

Q3. Write a note on method of Least Squares.

Answer :

The least square method is a statistical procedure which is used to find the best fit curve for the set of data where different variables are involved. This method is mostly used for the time series of data in which the relationship of two or more variables is difficult to identify. Least square method provides a trend line of best fit in the form of curve in order to represents the relationship between a known and unknown variable.

Q4. What do you mean by Deseasonalisation of Data?

Answer :

Deseasonalisation of data refers to the process of eliminating seasonal fluctuations from the given time series, it usually facilitates in adjusting the given time series for seasonal fluctuations and out with variables like trend component, cyclical and irregular variations.

Q5. Write about Seasonal Fluctuations.

Answer :

Seasonal fluctuations or variations refers to those periodic movements which occur regularly every year and have their origin in the nature of the year itself. They occur in regular and periodic manner over a span of less than a year, i.e., during a period of twelve months and have the same or almost the same pattern year after year.

UNIT 4

PROBABILITY

SYLLABUS

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.

LEARNING OBJECTIVES

✓ Concept of Probability with its Meaning and Importance.

✓ Basic Concepts of Probability – Experiment, Event, Mutually Exclusive Events, Collectively Exhaustive Events, Independent Events, Simple and Compound Events.

✓ Basics of Set Theory.

✓ Concept of Permutation and Combination.

✓ Approaches to Probability – Classical, Empirical, Subjective and Axiomatic

✓ Theorems of Probability – Addition and Multiplication Theorems

✓ Baye's Theorem and its Applications.

INTRODUCTION

Probability can be defined as the chance or 'likelihood of occurrence' of an experiment or event. Probability of any event ranges from 0 to 1.

A probability is a quantitative measure of uncertain events. It helps in determining the chances of the occurrence of an event. The systematic and extensive study of probability theory was made by 'B.Pascal', 'Pierre de Fermat', Jacques Bernoulli' in mid-seventeeth century.

The collection of finite or infinite number of objects with some common property is called Set. The objects belonging to the set are called Members or Elements of the set.

Permutations refer to different arrangements of objects in a set wherein all the elements are distinguishable. In these arrangements, an individual object is not repeated. If some objects are similar, then the permutations will be affected.

A combination is a selection of objects irrespective of their arrangement. The number of combinations of objects in different ways is entirely different from the number of their permutations.

Based on the concept of probability there are four different approaches of probability, which are explained below, Classical or priori approach, Relative frequency approach, Subjective approach, Axiomatic approach.

PART-A
SHORT QUESTIONS AND ANSWERS

Q1. Probability.

Answer :

Probability can be defined as the chance or 'likelihood of occurrence' of an experiment or event. Probability of any event ranges from 0 to 1. The term 'probability' was coined by an Italian mathematician, 'Galileo'. He was the first man who measured probability quantitatively while dealing with the problems associated with rolling of dice in gambling.

The word 'probability' or 'chance' is the most common word used in our day-to-day life. For example, in our daily life we use certain statements like, Probably he may win the elections etc.

A probability is a quantitative measure of uncertain events. It helps in determining the chances of the occurrence of an event. The systematic and extensive study of probability theory was made by 'B.Pascal', 'Pierre de Fermat', Jacques Bernoulli' in mid-seventeeth century.

Q2. What are mutually exclusive events, non-mutually exclusive events and dependent events?

OR

Explain (i) Mutually exclusive events and (ii) Not-mutually exclusive events.

(Refer Only Topics: Mutually Exclusive Events, Non-Mutually Exclusive Events)

May/June-18, Q5 (OU)

OR

Explain:

(i) Mutually exclusive events and

(ii) Dependent events.

(Refer Only Topics: Mutually Exclusive Events, Dependent Events)

Answer :

May/June-19, Q6 (OU)

Mutually Exclusive Events

Two events are said to be mutually exclusive (or incompatible) when both cannot happen simultaneously in a single trial. For example, if a coin is tossed, either a head or a tail appears but both cannot occur simultaneously.

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. For example, 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.

Dependent Events

Dependent events are the events in which the occurrence or non-occurrence of one event in any one trial influences the probability of occurrence of other events in other trials. For example, When a coin is tossed, getting a head or a tail are mutually exclusive and collectively exhaustive. Hence, if we get tail then head is considered as its complementary event.

Q3. Explain

(i) Dependent event and

(ii) Independent Event.

Answer : Jan.-21, Q6 (OU)

(i) Dependent Events

Dependent events are the events in which the occurrence or non-occurrence of one event in any one trial influences the probability of occurrence of other events in other trials.

Example

When a card is drawn from a pack of playing cards and is not replaced then this changes the probability of occurrence of the second card. Here, probability of occurrence of second event is dependent on the occurrence of first event.

(ii) Independent Events

Two or more events are considered as independent events, when the outcome of one event does not influences and is not influenced by the other event.

Example

When a student has appeared in physics and chemistry examinations, his marks obtained in physics is independent of the marks obtained in chemistry.

Q4. Explain the Axiomatic Approach to probability.

Answer : Jan.-21, Q7 (OU)

Axiomatic approach was given by the Russian mathematician, A.N.Kolmogorov in the year 1933 in his book "Foundation of Probability".

According to axiomatic approach, the probability of any event is calculated on basis of the axioms or postulates. Three axioms are taken into consideration for knowing the probability stated below,

(i) The probability of an event, if it does not occurs is 0(zero). For certain event it is 1(one) and for uncertain event it always ranges from 0 to 1.

(ii) The probability of the whole sample space is always $P(S) = 1$.

(iii) When 'A' and 'B' are two mutually exclusive events, then the probability of occurrence of either A or B is equal to the sum of their individual probabilities.

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

Q5. What do you mean by Conditional Probability?

Answer :

Two events 'A' and 'B' are said to be dependent events where 'B' occurs only when it 'A' knows that has been occurred or vice versa. Such a probability is called as 'Conditional Probability' for dependent events and is denoted by $P(A/B)$.

For statistically independent events, the conditional probability is same as that of the marginal probability.

For independent events,

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

For dependent events, $P(A/B) = P(A)$.

Q6. What is Joint Probability?

Answer :

A joint probability is the probability of occurrence of two or more simple events. It is the product of two marginal probabilities.

Joint probability of A and B is represented as $P(A \cap B)$

$$\therefore \quad P(A \cap B) = P(A).P(B)$$

Example

Probability of red queen picked from a pack of cards. It consists of two simple events of picking a red card and a queen.

Q7. What is Marginal Probability?

Answer :

Marginal probability is the simple probability of the occurrence of an event. It is also known as 'unconditional' probability or 'single probability'.

For example, when a single unbiased coin is tossed, the probability that it is a head is $P(H) = \frac{1}{2}$ and probability that it is a tail is $P(T) = \frac{1}{2}$.

These two events are independent and do not overlap one another i.e., the events are statistically independent of the outcomes of next coin been tossed.

The individual probabilities obtained in this case i.e., $P(H)$ or $P(T)$ are called as marginal probabilities.

Q8. When two dice are thrown, find the probability that the sum of the numbers is either 10 or 11.

Answer : (July/Aug-21, Q8 (MGU) | Sept/Oct-21, Q15 (OU) | May/June-18, Q6 (OU))

When two dice are thrown, the sample space, S contains $6^2 = 36$ points as shown below:

$$S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

The number of possible outcomes for getting the sums of 10 or 11 is,

(4, 6) (5, 5) (5, 6) (6, 4) (6, 5)

$\therefore$ The required probability is $\frac{5}{36}$.

Q9. 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?

Answer : May/June-19, Q7 (OU)

Let 'A' be the event of drawing a king from a standard pack of 52.

'B' be the event of drawing a queen.

$P(A)$ = Probability of drawing a king

$P(B)$ = Probability of drawing a queen.

As, the probability of drawing either a king or a queen has to be determined by drawing a single card, the events are said to be mutually exclusive.

For mutually exclusive events,

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

(By addition theorem of probability)

Given data,

$P(A) = \frac{4}{52}$ (As there are 4 kings in a standard pack of 52)

$P(B) = \frac{4}{52}$ (As there are 4 queens in a standard pack of 52)

$$\therefore P(A \cup B) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

Therefore, the probability of drawing either a king or a queen $= \frac{2}{13}$.

Q10. n(A) = 35, n(B) = 30, n(A ∩ B) = 20 then find ∩ (A ∪ B).

Answer : May/June-18, Q1(d) (KU)

Given that,

$$n(A) = 35$$
$$n(B) = 30$$
$$n(A \cap B) = 20$$
$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$
$$= 35 + 30 - 20$$
$$\therefore n(A \cup B) = 45$$

Q11. Find the number of ways in which the word ARRANGE can be arranged.

Answer : Sept./Oct.-21, Q4 (OU)

Given, the word ARRANGE

The total number of alphabets in ARRANGE are 7

The word ARRANGE can be arranged in 7 ways, i.e., $7p_7 = 7!$ Ways

The word ARRANGE consists of 2A's and 2R's, i.e., 2! and 2!

$$\therefore \text{Total Permutations} = \frac{n!}{(n-r)!}$$
$$= \frac{7!}{(7-5)! \times (7-5)!}$$
$$= \frac{7!}{2! \times 2!}$$
$$= \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1}$$
$$\therefore \text{Total Permutations} = 1{,}260$$

Q12. How many 3 letter words can be formed from the English world "SUCCESS"?

Answer : May/June-18, Q1(e) (KU)

The English word success formed by seven letters, which consists of,

$$S = 3 \text{ times}$$
$$C = 2 \text{ times}$$
$$U = 1 \text{ time and}$$
$$E = 1 \text{ time}$$

Number of ways for arranging these letters $= \frac{7!}{3!} = 840$ ways

Therefore, in the word "SUCCESS" 'S' is the letter which is repeating three times.

Q13. Find the value of $6P_4$, $5P_2$.

Answer : May/June-18, Q1(f) (KU)

(i) $6P_4$

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

Here, $n = 6$

$r = 4$

$$6P_4 = \frac{6!}{(6-4)!}$$

$$= \frac{6!}{2!}$$

$$= \frac{6\times5\times4\times3\times2\times1}{2\times1}$$

$$= 6 \times 5 \times 4 \times 3$$

$\therefore$ Value of $6P_4 = 360$

(ii) $5P_2$

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

Here,

$n = 5$

$r = 2$

$$5P_2 = \frac{5!}{(5-2)!}$$

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

$$= \frac{5\times4\times3\times2\times1}{3\times2\times1}$$

$$= 5 \times 4$$

$\therefore$ Value of $5P_2 = 20$

Q14. Calculate probability of 53 Mondays in a leap year.

Answer : May/June-19, Q4 (MGU)

A leap year consists of 366 days

A year has 52 weeks. Hence, there will be 52 mondays for sure.

52 weeks = 52 × 7 = 364 days

366 – 364 = 2 days

In a leap year, there will be 52 Mondays and 2 days will be left.

This 2 days may take the following combinations,

(Sunday, Monday), (Monday, Tuesday), (Tuesday, Wednesday), (Wednesday, Thursday), (Thursday, Friday), (Friday, Saturday), (Saturday, Sunday).

Out of 7 outcomes, favorable outcomes are 2.

Hence, the probability of getting 53 Mondays in a leap year is 2/7.

$\therefore$ It means that, there is a increase in prices over base level by 130.537%.

PART-B

ESSAY QUESTIONS AND ANSWERS

4.1 PROBABILITY – MEANING AND IMPORTANCE

Q15. What do you mean by probability? Explain the importance of probability.

Answer :

Probability

Probability can be defined as the chance or 'likelihood of occurrence' of an experiment or event. Probability of any event ranges from 0 to 1. The term 'probability' was coined by an Italian Mathematician, 'Galileo'. He was the first man who measured probability quantitatively while dealing with the problems associated with rolling of dice in gambling.

The word 'probability' or 'chance' is the most common word used in our day-to-day life. For example, in our daily life we use certain statements like 'Probably he may win the elections", "It is likely that India may win the match", "She may score above 90% in the upcoming examination" etc.

A probability is a quantitative measure of uncertain events. It helps in determining the chances of the occurrence of an event. The systematic and extensive study of probability theory was made by 'B.Pascal', 'Pierre de Fermat', Jacques Bernoulli' in mid-seventeeth century.

Importance of Probability

The various practical applications or importance of the theory of probability are as follows,

1. It helps in taking effective decisions under uncertain conditions.
2. It reduce complications in betting and games.
3. It is useful in economic decision making. It is very helpful in situations of risk and uncertainty.
4. It helps in carrying out different types of scientific investigations.
5. The fundamental laws of statistics, viz the law of statistical regularity and law of inertia of large numbers depends mainly on theory of probability.
6. It helps in decision theories which are based on fundamental laws of probability and expected value.
7. The empirical probability concept, based on experimental tests, provides scope for the application of probability to real life situations.
8. The subjective probabilities are used when it is not possible to do actual measurement. This has added a new dimension to the theory of probability. Such probabilities are revised by the application of Baye's rule.
9. It helps in making predictions with the help of probability models.
10. It helps in calculating probability values of different situations with the help of probability tree diagrams.

4.1.1 Experiment, Event, Mutually Exclusive Events, Collectively Exhaustive Events, Independent Events, Simple and Compound Events

Q16. Define probability. Explain the various concepts involved in probability.

Answer :

(Sept/Oct-21, Q16 (OU) | May/June-18, Q5(a) (KU))

Probability

Probability can be defined as the chance or 'likelihood of occurrence' of an experiment or event. Probability of any event ranges from 0 to 1. The term 'probability' was coined by an Italian Mathematician, 'Galileo'. He was the first man who measured probability quantitatively while dealing with the problems associated with rolling of dice in gambling.

Concepts of Probability

The key concepts or terms of probability are as follows,

1. Experiment

An experiment is also called as random experiment. It is a process or activity which leads to a particular outcome of several possible outcomes. The outcome which is going to be derived through random experiment is not known until its occurence (i.e., the outcome of random experiment is not predictable). But, the number of possible outcomes can be known. There may be fixed or infinite number of outcomes for a particular experiment. Outcome from random experiment may be numerical or non-numerical in nature.

Examples

Some of the examples of random experiment are as follows,

(i) Drawing a card from a group of shuffled cards.

(ii) Finding the chance of getting 6 when a dice is thrown.

2. Outcome

The result of a random experiment is usually referred as an outcome.

Example

If we toss a coin, the outcome may be a head or a tail. In such a case, number of outcomes = 2.

3. Event

An event is a possible outcome of an experiment or a result of trial. Basically there are two types of events. They are as follows,

(i) **Simple Event**

The probability of happening or non-happening of a single event is considered as a simple event.

Example

When we are selecting two black coins from a box containing 10 white and 5 black coins.

(ii) **Compound Event**

Compand event refers to joint occurrence of two or more events. It is also called as composite event'.

Example

A box containing 5 white balls, 3 black balls and 8 red balls, when we draw 2 white balls in first draw, 2 black balls in second draw and 5 red balls in third draw.

4. Mutually Exclusive Events

When two events cannot occur simultaneously in a single trial then such events are called as mutually exclusive or incompatible events.

Examples

(i) When a single coin is tossed, either a head or a tail can turn up, both cannot come at the same time.

(ii) At a single point of time, a person can be alive or dead.

5. 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.

Example

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. Collectively Exhaustive Events

Collectively exhaustive events are those events whose totality contains all the potential outcomes of a random experiment.

Example

When a dice is thrown, the total possible outcomes are 1, 2, 3, 4, 5 and 6 and thus the number of exhaustive cases is 6.

7. Equally Likely Events

Events are considered as equally likely events when the probability of occurrence of all the events are equal. These events are also called as 'equally, probable events'.

Example

If a coin is tossed, the two possible outcomes are head and tail. The probability of their occurrence is equal (i.e, $\frac{1}{2}$)

8. Independent Events

Two or more events are considered as independent events, when the outcome of one event does not influences and is not influenced by the other event.

Example

When a student has appeared in physics and chemistry examinations, his marks obtained in physics is independent of the marks obtained in chemistry.

9. **Dependent Events**

Dependent events are the events in which the occurrence or non-occurrence of one event in any one trial influences the probability of occurrence of other events in other trials.

Example

When a card is drawn from a pack of playing cards and is not replaced then this changes the probability of occurrence of the second card. Here, probability of occurrence of second event is dependent on the occurrence of first event.

10. **Complementary Events**

Two events are said to be complementary events if they are mutually exclusive and collectively exhaustive.

Example

When a coin is tossed, getting a head or a tail is mutually exclusive and collectively exhaustive. Thus, if we get tail then head is considered as its complementary event.

11. **Sample Space**

Sample space refers to the set of all possible outcomes for a particular random experiment. It is denoted by the capital letter S.

$$S = \{E_1, E_2, E_3, ... E_n\}$$

All the possible outcomes (sample space) can also be denoted by a tree diagram.

The main property of sample space is that, only one outcome from sample space can occur at a time (i.e.,) two or more outcome cannot occur simultaneously.

Example

In a experiment of tossing two coins, following sample events are the possible outcomes,

$E_1 = HH, E_2 = HT, E_3 = TH, E_4 = TT$

Here sample space, $S = \{E_1, E_2, E_3, E_4\}$

4.2 BASICS OF SET THEORY

Q17. Define set. How is it denoted and what are the two different ways of representing a set along with an example?

Answer :

Set

The collection of finite or infinite number of objects with some common property is called Set. The objects belonging to the set are called Members or Elements of the set.

Examples

(i) A group of people

(ii) A collection of stamps

(iii) A pair of glasses.

Notations

A set is usually denoted by capital letters with or without subscripts. Lower case letters are used to signify the elements of the set.

Representation of Set

There are two ways for representing a set. They are as follows,

1. **Roaster Method**

In Roaster Method, the elements are enclosed within the "{ }" brackets.

Example

Set of natural numbers

$N = \{1, 2, 3, 4, ...\}$

2. **Set-builder Form Method**

In Set-builder form method, the elements are described with respect to their common property.

Example

$\{x/x$ is a natural number$\}$

Note

If an element x belongs to set A, then it is represented as $x \in A$.

If an element x does not belong to set A, then it is represented as $x \notin A$.

Q18. Define the following with suitable example:

(i) Finite and infinite sets

(ii) Inclusion sets (Subset)

(iii) Equality of sets (Equal sets).

(iv) Proper subset

(v) Universal set

(vi) Null set

(vii) Singleton set

(viii) Power set

(ix) Disjoint set.

Answer :

(i) Finite and Infinite Sets

A set which contains finite (i.e., countable) number of elements, then that set is referred to as finite set. On the other hand, if a set contains infinite (i.e., uncountable) number of elements, then that set is referred as infinite set.

Example

{2, 4, 6, 8, ...} finite set

{1, 3, 5, 7, 9, ...} infinite set

Note

The order of finite set is denoted by $n(A)$ and the order of infinite set is denoted as ∞.

(ii) Inclusion Sets (Subset)

X and Y are two sets such that every element of X belongs to the set Y, then X is included in Y and X is called as subset of Y which is denoted by $X \subseteq Y$ or $Y \subseteq X$.

Example

$X_1 = \{1, 5, 7, 9\}$

$X_2 = \{1, 7\}$

X_2 is included in X_1

$X_2 \subseteq X_1$

Properties of Set Inclusion

For any three sets X, Y and Z

(a) Reflexive: $X \subseteq X$

(b) Transitive: $(X \subseteq Y)$ and $(Y \subseteq Z)$ then $X \subseteq Z$.

Note

Symmetric property is not satisfied (i.e., if $X \subseteq Y$, then it is not necessary that $Y \subseteq X$). Symmetric property is satisfied only when the sets are equal.

(iii) Equality of Sets (Equal Sets)

Two sets X and Y are said to be equal if $X \subseteq Y$ is equal to $Y \subseteq X$.

i.e., $X = Y \quad \Leftrightarrow (X \subseteq Y \wedge Y \subseteq X)$

Example

(a) Let $X = \{1, 2, 3, 4, 5\}$

$Y = \{1, 3, 5, 2, 4\}$

Here, $X \subseteq Y$ and $Y \subseteq X$ which imply $X = Y$.

(b) Let $X = \{\{2\}, \{3, 4\}, 5\}$

$Y = \{2, 3, 4, 5\}$

Here, $X \not\subseteq Y$ and $Y \not\subseteq X$ which imply $X \neq Y$.

Properties of Equality Sets

For any three sets X, Y and Z

(a) Reflexive $X = Y$

(b) Symmetric $X = Y \Rightarrow Y = X$

(c) Transitive $X = Y$, $Y = Z \Rightarrow X = Z$.

(iv) Proper Subset

For any set X if Y is the subset of X, but not equal to X, then Y is called proper subset of X. It is denoted by $Y \subset X$.

Symbolically $(Y \subseteq X \cap Y \neq X)$

Example

$X = \{1, 2, 3, 5, 7\}$

$Y = \{3, 5, 7\}$

$Y \subseteq X$

$Y \neq X \Rightarrow Y \subset X$.

(v) Universal Set

Any set which includes all the other sets defined is called universal set. It is denoted by μ or E.

Example

$A = \{x/x$ is a natural number$\}$

$B = \{x/x$ is a consonant$\}$

$C = \{x/x$ is an even prime number$\}$

$\mu = \{A, B, C\}$.

(vi) Null Set

A set with no element is called an empty or null set. It is denoted by ϕ or { }.

Note

For any set 'X', the null set 'ϕ' and the set X itself are subsets of set X.

(vii) Singleton Set

If a set consists of only one element then that set is referred as singleton set.

(viii) Power Set

If X is any set, then its power set consist of family of all its possible subsets. Power sets is denoted by $\rho^{(x)} = 2^X = \{x/x \subseteq x\}$.

Example

1. Let $X = \{1\}$

$\rho^{(x)} = \{\phi, \{1\}\} = \{\phi, X\}$

$\Rightarrow 2^1 = 2$

2. Let $Y = \{1, 2\}$

$\rho^{(x)} = \{\phi, \{1\}, \{2\}, \{1, 2\}\}$

$= \{\phi, \{1\}, \{2\}, S_2\}$

$\Rightarrow 2^2 = 4$

Note

If a set X contain n elements, then its power set contain 2^n elements.

(ix) Disjoint Set

If X and Y are two sets, which doesn't have any common elements (i.e., if $X \cap Y = \phi$) then these sets are said to be disjoint sets.

A collection of sets are said to be disjoint collection if there exist pair of disjoint sets. The elements belonging to disjoint collection are said to be mutually disjoint.

Q19. Explain with examples different operations that can be applied on the sets.

Answer :

The different operations performed on binary sets are as follows,

1. Intersection of Sets

X and Y are two sets, where the intersection of sets creates a new set that consists of the elements that are in X as well as in Y (i.e., common elements). Intersection of sets X and Y is denoted by $X \cap Y$.

Symbolically

$$A \cap B = \{x/(x \in A) \text{ and } (x \in B)\}$$

Properties of Intersection of Sets

(i) Commutative $X \cap Y = Y \cap X$

(ii) Associative $X \cap (Y \cap Z) = (X \cap Y) \cap Z$

(iii) $X \cap X = X$ (Idempotent)

(iv) $X \cap \phi = \phi$

2. Union of Sets

X and Y are two sets, where the union of set creates a new set that consist all elements present either in X or in Y in both. Union of sets is denoted by $X \cup Y$.

Symbolically

$$X \cup Y = \{x/(x \in X) \text{ or } (X \in Y)\}$$

Properties of Union of Sets

(i) Commutative: $X \cup Y = Y \cup X$

(ii) Associative: $X \cup (Y \cup Z) = (X \cup Y) \cup Z$

(iii) $X \cup X = X$ (Idempotent)

(iv) $X \cup \phi = X$.

Example

$$X = \{1, 2, 5, 7, 9\}$$
$$Y = \{3, 4, 5, 6, 8, 9, 7\}$$
$$X \cap Y = \{5, 7, 9\}$$
$$X \cup Y = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

3. Relative Complement (Difference)

X and Y are two sets, were relative complement of Y with respect to X, creates a new set that consist of elements belonging to X, but not Y. Relative complement is denoted by $X - Y$.

Symbolically

$$X - Y = \{x/(x \in X) \text{ and } (x \notin Y)\}$$

Properties of Relative Complement

(i) Commutative: $X - Y \neq Y - X$

(ii) Associative: $X - (Y - Z) \neq (X - Y) - Z$

Example

$$X = \{1, 2, 5, 7, 9\}$$
$$Y = \{3, 4, 5, 6, 8, 9, 7\}$$
$$X - Y = \{1, 2, 5, 7, 9\} - \{3, 4, 5, 6, 8, 9, 7\}$$
$$= \{1, 2\}$$
$$Y - X = \{3, 4, 6, 8\}$$

4. Symmetric Difference (Boolean Sum)

X and Y are two sets, where symmetric difference, creates a new set that consist of elements present either in A or B but not both.

It is denoted as $X + Y$ or $X \Delta Y = (X - Y) \cup (Y - X)$.

Symbolically

$$X \Delta Y = \{x/(x \in A) \text{ or } (x \in B) \text{ but not both}\}$$

Properties of Symmetric Difference

(i) Commutative: $X + Y = Y + X$

(ii) Associative: $X + (Y + Z) = (X + Y) + Z$

(iii) Idempotent: $X + X = \phi$

(iv) $X + \phi = X$.

Example

$$X = \{1, 2, 5, 7, 9\}$$
$$Y = \{3, 4, 5, 6, 8, 9, 7\}$$
$$X \Delta Y = (X - Y) \cup (Y - X)$$
$$= \{1, 2\} \cup \{3, 4, 6, 8\}$$
$$= \{1, 2, 3, 4, 6, 8\}$$

Absolute Complement

$\mu - A$ (i.e., relative complement of A with respect to μ) is called as absolute complement or complement of A.

PROBLEMS ON BASICS OF SET THEORY

Q20. In a college 500 students study Commerce and 400 students study Economics. If 300 students study both Commerce and Economics, find the total number of students enrolled into the college?

July/Aug.-21, Q1(d) (KU)

Solution :

Let A = The set of students studying Commerce

$\Rightarrow n(A) = 500$

B = The set of students studying Economics

$\Rightarrow n(B) = 400$

$n(A \cap B) = 300$ (studying both Commerce and Economics)

We need to find $n(A \cup B)$

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

$= 500 + 400 - 300 = 600$

∴ The total number of students enrolled in the college is 600.

Q21. 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?

Solution :

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 $\rho(B)$, 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 $A = \{1, 2, 3\}$, $B = \{1, 3, 6\}$

∴ $\rho(A) = \{0\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$

∴ $\rho(B) = \{0\}, \{1\}, \{3\}, \{6\}, \{1, 3\}, \{1, 6\}, \{3, 6\}, \{1, 3, 6\}$

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 $A \cup B$ 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

∴ 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.

$\rho(A)\cup\rho(B) = \{0\}, \{1\}, \{3\}, \{1, 3\}$ ...(3)

Now, $A\cap B = \{1, 3\}$

$\therefore \quad \rho(A\cap B) = \{0\}, \{1\}, \{3\}, \{1, 3\}$ ...(4)

From equations (3) and (4), we get,

$\rho(A)\cap\rho(B) \subseteq \rho(A\cap B)$

4.3 PERMUTATION AND COMBINATION

Q22. Explain in detail about permutations with examples.

Answer :

Permutations

Permutations refer to different arrangements of objects in a set where in all the elements are distinguishable. In these arrangements, an individual object is not repeated. If some objects are similar, then the permutations will be affected.

In other words, the different ways in which some objects (r) are selected and arranged from total number of objects (n) is called as "Permutations".

Example-1

If there are five different pens and three boxes, the number of arrangements of three pens in three boxes selected from five different pens is called as permutations.

The number of ways of arranging 'r' objects taken from 'n' different objects can be given by,

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

Where, '$n!$' is the number of ways of arranging 'n' objects.

$$n! = n(n-1)(n-2)(n-3)...$$

[$\because$ $n!$ is called as 'n' factorial, 0! is equal to 1]

For example,

$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$

From the above example, for calculating the different ways of arranging of three pens from 5 different pens in three different boxes,

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

Here,

$n = 5$

$r = 3$

$$\therefore \quad {}^5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!}$$

$$= \frac{5\times4\times3\times2\times1}{2\times1} = 60$$

$\therefore$ In 60 ways three pens selected from five different pens can be arranged in three different boxes.

Example-2

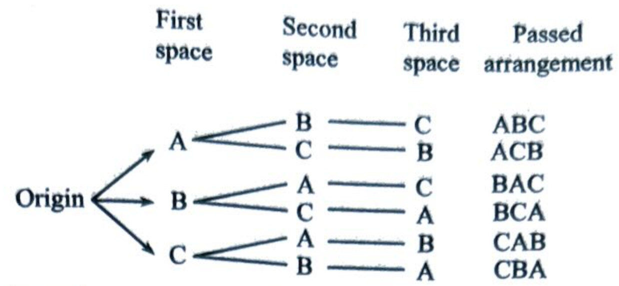
A factory owner has received three distinguishable new machines A, B and C and these can be arranged in 6 ways in the following ways,

$ABC, ACB, BAC, BCA, CAB, CBA$

From the above, it should be noted that each arrangement has three elements and no element appears twice

Example-3

The three distinguishable machines designated as A, B and C can be arranged on assembly line in the following manner.



The better way to see how three spaces can be filled by the three different machines is,

Spaces	1st	2nd	3rd
Ways	3	2	1

Applying the multiplication rule, the three spaces can be filled in,

$3 \times 2 \times 1 = 6$ ways, where repetition is not allowed.

Q23. Explain in detail about combination.

Answer :

Combination

A combination refers to a selection of objects irrespective of their arrangement. The number of combinations of objects in different ways is entirely different from the number of their permutations.

The total number of possible combinations of a set of objects is always taken as 1.

For example, the possible arrangements from the set of objects 'a' and 'b' are ab and ba. Irrespective of their order ab is same as ba, there is only one possible combination for this set.

The number of ways of selecting 'r' objects from 'n' different objects irrespective of their arrangement is called as combinations.

It is denoted by ^nC_r.

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

Proof

Each combination of 'r' objects which can be arranged in $r!$ ways gives rise to $r!$ permutations.

Hence, permutations of each of the ^nC_r combination yield ^nC_r permutations.

Thus, It can be represented as,

$$^nC_r \times r! = {^nP_r}$$

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

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

Note

In permutations we are concerned with the arrangement of objects whereas in combinations we are not concerned with the arrangement of objects. Thus, only one combination is possible.

For example,

Combination	Permutations
ABC	ABC,ACB,BAC,BCA,CAB,CBA

In the above example of three distinguishable machines namely A, B and C, there is only one combination i.e., ABC, whereas in permutations we find six different types of arrangement as shown in the above table.

PROBLEMS ON PERMUTATION AND COMBINATION

Q24. How many ways are there to paste 2 photos on notice board from a group of 6 photos?

May/June-18, Q5(b) (KU)

Solution :

Given that,

Total Photos, $n = 6$

Photos to paste on the notice board, $r = 2$

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

$$6P_r = \frac{6!}{(6-2)!} = \frac{6!}{4!}$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1}$$

$$= 6 \times 5 = 30$$

$\therefore$ 30 ways are there to paste 2 photos on notice board from a group of 6 photos.

Q25. Find the probability of getting 3 heads when 6 coins are tossed simultaneously?

July/Aug.-21, Q9 (MGU)

Answer :

Given that,

Number of tosses, n = 6

Probability of getting a head, $p = \frac{1}{2}$

Probability of not getting a head, $q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

Probability of getting 3 heads $= {}^nC_r \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{n-r}$

$$= {}^6C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{6-3}$$

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

$$= \frac{6!}{3! \times 3!} \times \left(\frac{1}{2}\right)^{3+3}$$

$$= \frac{{}^2\cancel{6} \times 5 \times \cancel{4}^2 \times 3!}{3! \times \cancel{3}_1 \times \cancel{2}_1 \times 1} \times \left(\frac{1}{2}\right)^6 = 20 \times \left(\frac{1}{2}\right)^6$$

$$= 20 \times \frac{1}{65} = \frac{5}{16}$$

4.4 APPROACHES TO PROBABILTY: CLASSICAL, EMPIRICAL, SUBJECTIVE, AXIOMATIC

Q26. Define probability. Explain any two approaches to probability.

July/Aug.-21, Q8 (KU)

OR

What are the various theories or approaches used in probability?

(Refer Only Topic: Approaches to Probability)

Answer :

Probability

For answer refer Unit-IV, Page, No. 102, QNo. 15, Topic: Probability.

Approaches to probability

Based on the concept of probability, there are four different approaches of probability. They are as follows,

1. Classical or Priori Approach

Classical or priori approach is given by, "James Bernoulli". He was the first person who found the uncertainty quantitatively.

The classical or priori approach assumes that all the outcomes of a random experiments are mutually exclusive, and equally likely. According to priori approach in a random experiment when there are 'm' favourable cases to a favourable, then event 'A' occurs. If 'n' has total possible cases, then favourable event 'A' does not occur. Thus, the probability of getting favourable cases $P(A)$ can be calculated as,

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

Importance of classical theory

The importance of classical theory are as follows,

(i) The probability of an event always lies between 0 and 1.

$$0 \leq P(A) \leq 1$$

(ii) The sum of probability of an event and its complement is always 1.

$$p + q = 1$$

(iii) The sum of the probability of success and failure is 1.

For example,

'p' is the probability of getting a head and

'q' is the probability of getting a tail.

Then,

$$p + q = \frac{1}{2} + \frac{1}{2} = 1$$

(iv) The probability of occurrence for any event has 3 chances.

(a) In case of certain event, $P(E) = 1$

(b) In case of impossible event, $P(E) = 0$

(c) In case of uncertain event, $0 \leq P(E) \leq 1$.

The limitations of classical theory are as follows,

(i) The outcomes are not equally likely.

(ii) The collectively exhaustive events of an experiment are infinite.

(iii) The classical theory does not provide answers to certain question which occurs in our daily life.

For example,

(a) What is the probability of occurrence of rain now?

(b) The chances of bulb getting failed etc.

2. Empirical or Relative Frequency Approach

Relative frequency approach was given by Richard Von Mises. In some cases, the desired event may or may not occur. This approach assumes that a random experiment is repeated number of times under identical conditions and the trials are independent to each other.

Relative frequency approach calculates the proportion of time known as relative frequency with which the event takes place repeatedly over an infinite number of times under identical conditions.

Therefore, the probability of occurring an event A, can be calculated as,

$$P(A) = \lim_{n \to \infty} \left\{ \frac{a}{n} \right\}$$

Where,

'a' is the number of times an event 'A' is repeated

'n' is the trials of an experiment.

As the probability of an event is ascertained by repetitive empirical observations, this probability is known as 'empirical probability'.

Some of the limitations of this approach are,

(i) It takes large amount of time, as the experiments are repeated large number of times

(ii) During the experimental time, the conditions may not be always identical and homogenous.

3. Subjective Approach

Subjective approach was initiated by "Frank Ramsey" in the year 1926 in his book, "The Foundation of Mathematics and Other Logical Essays".

Subjective approach depends on the extent to which the experiences focusses on the chances of the occurrence of random event. Hence, it is also known as 'personalistic approach'.

Some of the limitations of this approach are,

(i) As the probability is just an estimation based on one's personal beliefs it may differ from one person to another person.

(ii) It cannot calculate probability accurately thus the results are not known.

(iii) There is no particular formula for calculating probability quantitatively.

4. Axiomatic Approach

Axiomatic approach was given by the Russian mathematician, A.N.Kolmogorov in the year 1933 in his book "Foundation of Probability".

According to axiomatic approach, the probability of any event is calculated on basis of the axioms or postulates. Three axioms are taken into consideration for knowing the probability stated below,

(i) The probability of an event, if it does not occurs is 0(zero). For certain event it is 1(one) and for uncertain event it always ranges from 0 to 1.

(ii) The probability of the whole sample space is always $P(S) = 1$.

(iii) When 'A' and 'B' are two mutually exclusive events, then the probability of occurrence of either A or B is equal to the sum of their individual probabilities.

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

PROBLEMS ON BASIC PROBABILITY

Q27. Two dice are rolled. Find the probability of 6 number event when two dice are rolled.

Solution : May/June-19, Q9(b) (MGU)

When two dice are rolled, the sample space, S contains $6^2 = 36$ outcomes as shown below,

$$S = \begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{Bmatrix}$$

Let 'E' be the event of getting 6.

Possible number of outcomes for the event 'E' is n(E) = 5

$$\text{Probability of event 'E'} \Rightarrow p(E) = \frac{n(E)}{n(S)} = \frac{5}{36}$$

$\therefore$ Probability of 6 number event when two dice are rolled is $\frac{5}{36}$.

Q28. 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.

Jan.-21, Q15 (OU)

Solution :

(i) When There is Replacement

Given,

Total number of balls in a box = 8 + 5 = 13

3 balls can be drawn from 13 in $^{13}C_3$ Ways.

3 red balls can be drawn form 8 in $^{8}C_3$ Ways

3 White balls can be drawn from 5 in $^{5}C_3$ Ways.

The probability of drawing 3 White balls in the first trial can be find out in the following manner,

$P(AnB) = P(A).P(B/A)$

Let A is the event such that the first drawing will give 3 White balls.

Let B is the event such that the Second drawing will given 3 red balls.

Therefore,

$$P(AnB) = P(A).P(B/A)$$

The probability of 3 white balls at first trial is,

$$P(A) = \frac{^5C_3}{^{13}C_3} = \frac{5}{143}$$

The probability of 3 red balls at the second trial is,

$$P(B/A) = \frac{8c_3}{13c_3} = \frac{28}{143}$$

$$\therefore P(AnB) = \frac{5}{143} \times \frac{28}{143} = \frac{140}{20,449}$$

$$\Rightarrow 0.007$$

Working Notes (WN)

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

1.

$$^5C_3 = \frac{5!}{3!(5-3)!}$$

$$= \frac{5\times4\times3\times2\times1}{3\times2\times1(2)!}$$

$$= \frac{5\times4\times\cancel{3\times2\times1}}{\cancel{3\times2\times1}(2\times1)}$$

$$= \frac{5\times4}{2\times1}$$

$$= \frac{20}{2} = 10$$

2.

$$^{13}C_3 = \frac{13!}{3!(13-3)!}$$

$$= \frac{13\times12\times11\times\cancel{10\times9\times8\times7\times6\times5\times4\times3\times2\times1}}{3\times2\times1(\cancel{10\times9\times8\times7\times6\times5\times4\times3\times2\times1})}$$

$$= \frac{13\times\cancel{12}^4\times11}{\cancel{3}\times2\times1}$$

$$= \frac{13\times4\times11}{1\times2\times1}$$

$$= \frac{572}{2} = 286$$

$$\therefore \frac{8c_3}{13c_3} = \frac{\cancel{10}^5}{\cancel{28}_{143}} = \frac{5}{143}$$

(ii) When There is No Replacement

The probability of drawing 3 white balls in the first trial is,

$$P(AnB) = P(A).P(B/A)$$

$$P(A) = \frac{^5C_3}{^{13}C_3} = \frac{5}{143}$$

When the white balls are drawn and are not replaced, the box contains 2 white balls and 8 red balls.

$\therefore$ At, the second trial 3 balls can be drawn from 10 in $^{10}C_3$ Ways and 3 red balls can be drawn from 8 in $^{8}C_3$ Ways.

The probability of 3 red balls in the second trial,

$$P(B/A) = \frac{^7C_3}{^{10}C_3} = \frac{7}{24}$$

$$\therefore P(AnB) = \frac{5}{143} \times \frac{7}{24} = \frac{7}{429}$$

$$= 0.0102.$$

Working Notes (WN)

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

1.

$$^7C_3 = \frac{7!}{3!(7-3)!}$$

$$= \frac{7\times6\times5\times4\times3\times2\times1}{3\times2\times1(4)}$$

$$= \frac{7\times6\times5\times\cancel{4\times3\times2\times1}}{3\times2\times1(\cancel{4\times3\times2\times1})}$$

$$= \frac{\cancel{210}^{35}}{\cancel{6}_1} = 35$$

2.

$$^{10}C_3 = \frac{10!}{3!(10-3)!}$$

$$= \frac{10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{3\times2\times1(7)!}$$

$$= \frac{10\times9\times8\times\cancel{7\times6\times5\times4\times3\times2\times1}}{3\times2\times1(\cancel{7\times6\times5\times4\times3\times2\times1})}$$

$$= \frac{10\times\cancel{9}^3\times\cancel{8}^4}{\cancel{3}_1\times\cancel{2}_1\times1} = \frac{720}{6}$$

$$= 120$$

$$\frac{^7C_3}{10C_3} = \frac{35}{120} = \frac{7}{24}$$

Q29. In a group of 60 students, 35 students can speak Hindi, while 25 can speak both Hindi and Telugu. All students can speak at least one language. How many students can speak Hindi but not Telugu?

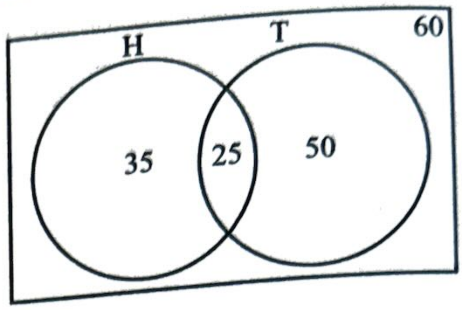
July/Aug.-21, Q7 (MGU)

Solution :

Let A denote the set of people who speak Hindi and B the set of people who speak Telugu.

Then, n (A) = 35, n (B) = ?, n (A∪B) = 60

n (A∩B) = 25



n (A∪B) = n (A) + n (B) − n (A∩B)

60 = 35 + n (B) − 25

= n (B) = 60 − 35 + 25 = n (B) = 50

n (A − B) = A (A) − n (A∩B)

= 35 − 25

n (A − B) = 10

∴ Number of students who speak Hindi but not Telugu = 10.

4.5 THEOREMS OF PROBABILITY: ADDITION, MULTIPLICATION

Q30. Explain the probability theorem basic concepts

May/June-19, Q9(a) (MGU)

OR

Explain the major theorems of probability.

Answer :

The major and important theorems of probability are as follows,

1. Addition Theorem

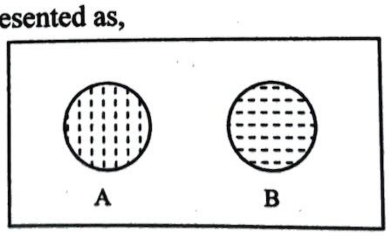
Addition theorem is different for mutually exclusive events and non-mutually exclusive events.

For Mutually Exclusive Events

When 'A' and 'B' are two mutually exclusive events (i.e., both cannot occur at the same time) then the probability 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,

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

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,

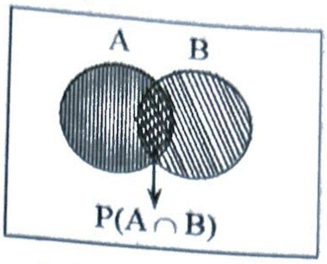


Figure: Non-Mutually Exclusive Events

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,

$$P(A \text{ or } B \text{ or } C) = P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

2. **Multiplication Theorem**

If 'A' and 'B' are two independent events then the probability occurrence of both the events is equal to the product of their individual probabilities.

For independent events,

$$P(A \cap B) = P(A).P(B)$$

Similarly,

$$P(A \cap B \cap C) = P(A).P(B).P(C)$$ 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,

$$P(A \cap B) = P(A/B).P(B)$$
$$= P(B/A).P(A)$$

Where, $P(A/B)$ is a conditional probability of A given that B has occurred (The probability of occurrence of event A when event B has already occurred is the conditional probability of A given B).

PROBLEMS ON THEOREMS OF PROBABILITY

Q31. 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.

Solution :

May/June-19, Q12(a) (OU)

Given that,

A Bag contains 4 defective and 6 good electronic calculators,

⇒ Total number of calculators = 10

(i) **Probability That Two are Good**

Probability of getting good calculator in first draw is $\frac{6}{10} = \frac{3}{5}$

After one calculator is drawn, 9 calculators are left in the bag as the first one is not replaced. Now the probability of getting good calculator in second draw is $\frac{5}{9}$

∴ The probability of getting two good calculator is $\left(\frac{3}{5} \times \frac{5}{9}\right) = \frac{1}{3}$

(ii) Probability That Two are Defective

Probability of getting defective calculator in first draw is $\frac{4}{10} = \frac{2}{5}$

After one calculator is drawn, only 9 calculators are left in the bag. Now, the probability of getting defective calculator in second draw is $\frac{3}{9} = \frac{1}{3}$

$\therefore$ The probability of getting two defective calculator is $\frac{2}{5} \times \frac{1}{3} = \frac{2}{15}$

(iii) Probability That One is Good and One is Defective

Case (i)

First one is good and second one is defective

$$\text{Probability} = \frac{6}{10} \times \frac{4}{9} = \frac{4}{15}$$

Case (ii)

First one is defective and second one is good

$$\text{Probability} = \frac{4}{10} \times \frac{6}{9} = \frac{4}{15}$$

Probability of union of these two alternative cases is their sum

$$= \frac{4}{15} + \frac{4}{15} = \frac{8}{15}.$$

Q32. The probability that a contractor will get a plumbing contract is 3/4 and the probability that he will not get electric contract is 4/9. If the probability of getting at least one contract is 5/6, what is the probability that he will get both the contracts? Use addition theorem to solve this problem.

Solution :

Let 'A' be the event that the contractor will get plumbing contract.

'B' be the event that the contractor will get electric contract.

The probability of getting plumbing contract = $P(A)$

The probability of not getting a plumbing contract = $P(\overline{A})$

The probability of getting electric contract = $P(B)$

The probability of not getting electric contract = $P(\overline{B})$

The probability of getting at least one contract is,

$$P(A \text{ or } B) = \frac{5}{6}$$

Given that,

$$P(A) = \frac{3}{4};\ P(\overline{A}) = \frac{1}{4} \quad (\because P(A) + P(\overline{A}) = 1)$$

$$P(\overline{B}) = \frac{4}{9};\ P(B) = \frac{5}{9}$$

$\therefore$ The probability that he will get both the contracts can be given by using addition theorem.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A \text{ and } B) = P(A) + P(B) - P(A \text{ or } B)$$

$$= \frac{3}{4} + \frac{5}{9} - \frac{5}{6}$$

$$= \frac{27+20-30}{36} = \frac{47-30}{36}$$

$$= \frac{17}{36} = 0.472 = 47.2\%$$

∴ The probability that the contractor will get both the contracts = 47.2%

4.6 BAYE'S THEOREM

Q33. State and explain Baye's Probability theorem with its applications.

Answer :

Baye's Theorem

Baye's theorem got its name from the British mathematician, 'Thomas Bayes' in 1763. Baye's theorem deals with revising of priori probability by making use of new information for calculating posterior probabilities. It is also known as 'Rule for the Inverse Probability'.

In other words, priori probabilities are revised or converted into posterior probabilities (or revised probabilities) by using new information in Baye's theorem.

The probabilities of an event before the collection of new information is known as 'Priori probabilities'. Where as, Posterior probabilities are the revised priori probabilities that have been derived after using the new information. It is also known as 'Inverse' or 'Revised' probabilities.

If, $A_1, A_2, A_3 \ldots A_n$ are mutually exclusive and collectively exhaustive events, then

$P(A_1), P(A_2), P(A_3), \ldots, P(A_n)$ are the priori probability,

'B' is an event such that $P(B) \neq 0$ whose conditional probabilities are represented as,

$$P(B/A_1), P(B/A_2), P(B/A_3), \ldots\ldots, P(B/A_n)$$

It should be noted that the conditional probabilities are known with the help of this data we have to calculate posterior probabilities.

∴ Posterior probabilities can be calculated by using the following formula,

$$P(A_i / B) = \frac{P(A_i \cap B)}{\sum_{i=1}^{n} (A_i \cap B)} = \frac{P(B / A_i)P(A_i)}{\sum_{i=1}^{n} P(B/A_i)\, P(A_i)}$$

(or)

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

Where,

$P(A_i \cap B)$ is the joint probability of A_i and B events

A_i and B events

$P(A_i)$ is the priori probability and

$P(B/A_i)$ is the conditional probability.

$$\therefore\ P(B) = \sum_{i=1}^{n} P(B / A_i)\, P(A_i)$$

Applications of Baye's Theorem

The following points highlights the application of Baye's theorem,

1. The posterior probabilities can be known by revising priori probabilities with the help of new information.
2. The probability of occurrence of future events can be known by Baye's theorem.
3. The powerful statistical tools are being offered.
4. The baye's theorem helps the business and management executives to take effective decisions in uncertain situations.

Baye's theorem is also known as 'Probability of Causes' as it helps in determining the probability where a particular effect has due to a specific cause.

The entire revision process of priori probabilities can be diagrammatically represented as follows,

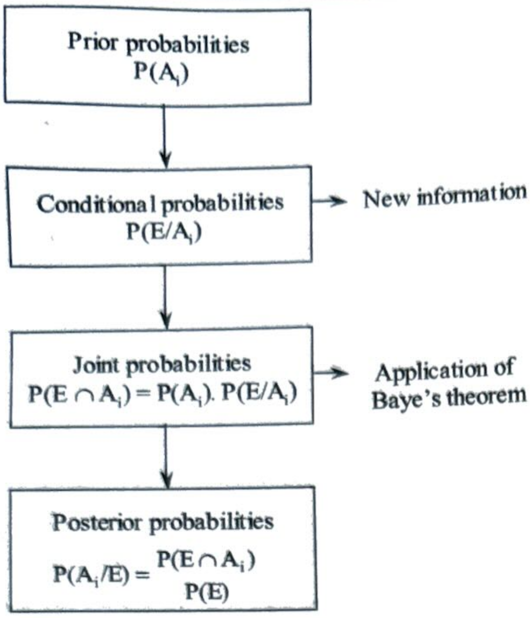


Figure: Priori Probabilities Process

PROBLEMS ON BAYE'S THEOREM

Q34. A company has two Plants for manufacturing Scooters. Plant I manufactures 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.

Solution : May/June-19, Q12(b) (OU)

(i) It is manufactured by Plant I

Let,

The scooter produced by plant-I = A_1

The scooter produced by plant-II = A_2

The standard quality scooter produced by plant I and II = E

$$P(A_1) = 80\% \text{ or } 0.80$$

$$P(A_2) = 20\% \text{ or } 0.20$$

$$P(E/A_1) = \frac{85}{100} = 0.85$$

$$P(E/A_2) = \frac{65}{100} = 0.65$$

$$P(E \cap A_1) = P\left(\frac{E}{A_1}\right) \times P(A_1)$$

$$= 0.85 \times 0.80 = 0.68$$

$$P(E \cap A_2) = P\left(\frac{E}{A_2}\right) \times P(A_2)$$

$$= 0.65 \times 0.20 = 0.13$$

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

As per Baye's theorem the probability to manufacture standard quality scooter chosen at random

$$= P\left(\frac{A_1}{E}\right)$$

$$= \frac{P\left(\frac{E}{A_1}\right) \times P(A_1)}{P\left(\frac{E}{A_1}\right) \times P(A_1) + P\left(\frac{E}{A_2}\right) \times P(A_2)}$$

$$= \frac{0.68}{0.68 + 0.13}$$

$$= \frac{0.68}{0.81}$$

$$= 0.839 \text{ (or) } 0.84$$

$$= 0.84\% \text{ or } \frac{68\%}{81\%}$$

$$P\left(\frac{A_2}{E}\right) = \frac{P\left(\frac{E}{A_2}\right) \times P(A_2)}{P\left(\frac{E}{A_2}\right) \times P(A_2) + P\left(\frac{E}{A_1}\right) \times P(A_1)}$$

$$= \frac{0.13}{0.13 + 0.68} = \frac{0.13}{0.81}$$

$$= 0.16$$

$$= 0.16\% \text{ or } \frac{13\%}{81\%}$$

Q35. 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?

Solution :

Jan.-21, Q16 (OU)

Let 'A_1' be the event of drawing an item produced by machine 1.

While, 'A_2' is the event of drawing an item produced by machine 2.

Let, 'B' be the event of drawing defective item which is produced by either of the machines. From the additional information given,

$P(A_1)$ 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 item produced by machine 1.

$P(B/A_2)$ is the probability of getting defective item produced by machine 2.

From the given data,

$P(A_1) = 0.3$ (30%) $P(B/A_1) = 5\% = \frac{5}{100} = 0.05$

$P(A_2) = 0.7$ (70%) $P(B/A_2) = 1\% = \frac{1}{100} = 0.01$

Computation of Posterior Probabilities

Events	Priori Probability $P(A_i)$	Conditional Probability $P(B/A_i)$	Joint Probability $P(A_i \cap B)$	Posterior Probability $P(A_i/B) = \frac{P(A_i \cap B)}{\Sigma(A_i \cap B)}$
A_1	$P(A_1) = 0.3$	$P(B/A_1) = 0.05$	$0.3 \times 0.05 = 0.015$	$P(A_1/B) = \frac{0.015}{0.022} = 0.682$
A_2	$P(A_2) = 0.7$	$P(B/A_2) = 0.01$	$0.7 \times 0.01 = 0.007$	$P(A_2/B) = \frac{0.007}{0.022} = 0.318$
Total	**1.00**	**0.06**	$\Sigma P(A_i \cap B) = 0.022$	**1.000**

Therefore, the posterior probability obtained after calculation are as follows,

If a defective item is drawn at random,

(i) The probability of defective item produced by machine 1, $P(A_1/B) = 0.682$ or $= 68\%$

(ii) The probability of defective item produced by machine 2, $P(A_2/B) = 0.318 = 31.8\%$ or $\simeq 38$

Note:

Always the sum of priori probability as well as posterior probabilities is equal to 1 or 100.

Q36. Assume that a factory has two machines. Past records show that machine 1 produces 40% of the items of output and machine 2 produces 60% of the items. Further, 3% of the items produces by machine 1 were defective and only 1% produced by machine 2 were defective. If a defective item is drawn at random, what is the probability that the defective item was produced by machine 1 or machine 2.

Solution :

July/Aug-21, Q9 (KU)

Let 'A_1' be the event of drawing an item produced by machine 1.

While, 'A_2' is the event of drawing an item produced by machine 2.

Let, 'B' be the event of drawing defective item which is produced by either of the machines. From the additional information given,

$P(A_1)$ 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 item produced by machine 1.

$P(B/A_2)$ is the probability of getting defective item produced by machine 2.

From the given data,

$P(A_1) = 0.4\ (40\%)\ P(B/A_1) = 3\% = \frac{3}{100} = 0.03$

$P(A_2) = 0.6\ (60\%)\ P(B/A_2) = 1\% = \frac{1}{100} = 0.01$

Computation of Posterior Probabilities

Events	Priori Probability $P(A_i)$	Conditional Probability $P(B/A_i)$	Joint Probability $P(A_i \cap B)$	Posterior Probability $P(A_i/B) = \frac{P(A_i \cap B)}{\Sigma(A_i \cap B)}$
A_1	$P(A_1) = 0.4$	$P(B/A_1) = 0.03$	$0.4 \times 0.03 = 0.012$	$P(A_1/B) = \frac{0.012}{0.018} = 0.67$
A_2	$P(A_2) = 0.6$	$P(B/A_2) = 0.01$	$0.6 \times 0.01 = 0.006$	$P(A_2/B) = \frac{0.006}{0.018} = 0.33$
Total	**1.00**	**0.04**	$\Sigma P(A_i \cap B) = 0.018$	**1.000**

Therefore, the posterior probability obtained after calculation are as follows,

If a defective item is drawn at random,

(i) The probability of defective item produced by machine 1, $P(A_1/B) = 0.67$ or $= 67\%$

(ii) The probability of defective item produced by machine 2, $P(A_2/B) = 0.33$ or $= 33\%$

Note:

Always the sum of priori probability as well as posterior probabilities is equal to 1 or 100.

Q37. 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.

Solution : **May/June-18, Q12(b) (OU)**

Let P (A) be the probability of the event that the bolt is manufactured by the machine P.

P(B) be the probability of the even t that the bolt is manufactured by the machine Q.

P(C) be the probability of the event that the bolt is manufactured by the machine R.

Probability of manufactured of bolt factory machines 'P'

$P(A) = 25\% = 0.25$ (given)

Probability of manufactured bolt factory machines 'Q'.

$P(B) = 35\% = 0.35$ (given)

Probability of manufactured bolt factory machines (R)

$P(C) = 40\% = 0.4$ (given)

Let 'D' be the number the defective items produced by bolt factory.

∴ The probability of defective items produced by A,

$P(D/A) = 0.05$ (given)

Probability of defective event manufactured by B,

$P(D/B) = 0.04$

Probability of defective event manufactured by C,

$P(D/C) = 0.02$

Now calculating the joint probabilities,

Event A, $P(A \cap D) = P(A) \times P(D/A)$

$= 0.05 \times 5\ 0.25$

Event B, $P(B \cap D) = P(B) \times P(D/B)$

$= 0.04 \times 4 = 0.16$

Event C, $P(C \cap D) = P(C) \times P(D/C)$

$= 0.02 \times 2 = 0.04$

For getting the required probabilities, Baye's theorem is applied in this case.

Computing the values by using Baye's theorem is tabulated as follows,

Event (1)	Probability (2)	Conditional Probability (3)	Joint Probability = (2) × (3) (4)	Posterior Probability = $\frac{(4)}{0.45}$ (5)
A	$P(A) = 0.05$	$P(D/A) = 5$	$P(A \cap D) = 0.05 \times 5 = 0.25$	$P(A/D) = \frac{0.25}{0.45} = 0.55$
B	$P(B) = 0.04$	$P(D/B) = 4$	$P(B \cap D) = 0.04 \times 4 = 0.16$	$P(B/D) = \frac{0.16}{0.45} = 0.35$
C	$P(C) = 0.02$	$P(D/C) = 2$	$P(C \cap D) = 0.02 \times 2 = 0.04$	$P(C/D) = \frac{0.04}{0.45} = 0.8$
Total	0.11		0.45	0.98

∴ The probability of a defective event drawn at random is,

From machine P = 0.55 = 55%

From machine Q = 0.35 = 35%

From machine R = 0.08 = 8%

EXERCISE AND PRACTICE QUESTIONS

SHORT QUESTIONS

THEORY

Q1. Define Probability. [Refer, Q1]

Q2. What are mutually exclusive events, non-mutually exclusive events and dependent events?

OR

Explain (i) Mutually exclusive events and (ii) Not-mutually exclusive events.

May/June-18, Q5 (OU)

OR

Explain:

(i) Mutually exclusive events and

(ii) Dependent events. [Refer, Q2] May/June-19, Q6 (OU)

Q3. What do you mean by Conditional Probability? [Refer, Q5]

Q4. What is Joint Probability? [Refer, Q6]

Q5. What is Marginal Probability? [Refer, Q7]

PROBLEMS

Q6. $n(A) = 70$, $n(B) = 60$, $n(A \cap B) = 40$ then find $\cap (A \cup B)$. [Refer Similar, Q10]

(Ans: $n(A \cup B) = 90$).

Q7. How many 5 letter words can be formed from the English world "ACCESS"? [Refer Similar, Q12]

Q8. Find the value of $8P_6$, $6P_4$. [Refer Similar, Q13]

Q9. Calculate probability of 53 Tuesdays in a leap year. [Refer Similar, Q14]

ESSAY QUESTIONS

THEORY

Q1. What do you mean by probability? Explain the importance of probability. [Refer, Q15]

Q2. Define probability. Explain the various concepts involved in probability. [Refer, Q16]

(Sept/Oct-21, Q16 (OU) | May/June-18, Q5(a) (KU))

Q3. Define set. How is it denoted and what are the two different ways of representing a set along with an example? [Refer, Q17]

Q4. Explain with examples different operations that can be applied on the sets. [Refer, Q19]

Q5. Explain in detail about permutations with example. [Refer, Q22]

Q6. What are the various theories or approaches used in probability? [Refer, Q26]

Q7. Explain the probability theorem basic concepts. [Refer, Q30] May/June-19, Q9(a) (MGU)

Q8. State and explain Baye's Probability theorem with its applications. [Refer, Q33]

PROBLEMS

Q9. In a certain class there are 21 students in subject X, 17 in subject Y and 10 in subject Z. Of these 12 attend subjects X and Y, 5 attend subjects Y and Z, 6 attend subject X and Z. These include 2 students who attend all the three subjects. Find the probability that a student studies one subject alone. [Refer Similar Q36]

(Ans: 8/27).

Q10. A box contains 4 red pens and 5 black pens. Find the probability of drawing 3 black pens one by one, [Refer Similar Q28]

(i) With replacement (ii) Without replacement.

(Ans: (i) 125/729, (ii) 5/42).

Q11. Consider a standard deck of cards. One card is drawn at random. Find the probability of drawing either a king or a red card. [Refer Similar Q9]

(Ans: 7/13).

Q12. A manufacturing firm produces steel pipes in three plants with daily production volumes of 500, 1000 and 2000 units respectively. According to past experience, it is known that the fraction of defective outputs produced by the three plants are respectively 0.005, 0.008, 0.010. If a pipe is selected from the day's total production and found to be defective, find out,

From which plant the pipe comes? (or)

What is the probability that it comes from the first plant, 2nd plant and third plant? [Refer Similar Q36]

(Ans: Total Production = 3,500 Units; Plant A = 8.27%, Plant B = 26.18%, Plant C = 65.55%).

Q13. 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? [Refer Similar Q35]

(Ans: A = 0.6, B = 0.4).

Q14. If A and B are two sets, then define $A \triangle B = (A - B) \cup (B - A)$. If A = {1, 2, 3} B = {1, 3, 5} then find the set

$((A \triangle B) \triangle B) - (A \triangle (B \triangle B))$. [Refer Similar Q21]

(Ans: $(A \triangle B) \triangle B = \{1, 2, 3\}$, $A \triangle (B \triangle B) = \{1, 2, 3, \phi\}$).

Q15. If A = {1, 2, 3} B = {2, 3, 4} C = {1, 3, 4} then find, [Refer Similar Q21]

(i) $A - (B - C)$ (ii) $A - (B \cap C)$

(iii) $(A \cup B) - C$.

(Ans: $A - (B - C) = \{1, 3\}$, $A - (B \cap C) = \{1, 2\}$; $(A \cap B) - C = \{2\}$).

Q16. 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. [Refer Similar Q31]

(Ans: The probability of having a multiple of 3 or 7 is $\frac{13}{30}$) May/June-18, Q12(a) (OU)

INTERNAL ASSESSMENT/EXAM

I Multiple Choice

1. The term 'probability' was first coined by an Italian Mathematician, ________. [ ]
   (a) Flintof (b) Steven Smith
   (c) Galileo (d) Gayle

2. ________ is also known as 'mathematical probability'. [ ]
   (a) Classical probability (b) Statistical probability
   (c) Axiomatic probability (d) Chain probability

3. The collection of finite or infinite number of objects with some common property is called as ________. [ ]
   (a) Set (b) Pair
   (c) Probability (d) Method

4. In ________ the elements are enclosed with in the "{ }" brackets. [ ]
   (a) Cutter method (b) Disaster method
   (c) Roaster method (d) Toaster method

5. ________ is different for mutually exclusive events and non-mutually exclusive events. [ ]
   (a) Addition theorem (b) Multiplicative theorem
   (c) Division theorem (d) Substractive theorem

6. Baye's theorem got its name from the British mathematician, ________. [ ]
   (a) Andrew Bayes (b) Ricky Bayes
   (c) Chris Bayes (d) Thomas Bayes

7. Priori probabilities are revised or converted into ________ probabilities. [ ]
   (a) Interior (b) Posterior
   (c) Exterior (d) In-exterior

8. ________ theorem offers a powerful statistical tool. [ ]
   (a) Ross (b) Jos
   (c) Baye's (d) Gross

9. Baye's theorem is also known as ________. [ ]
   (a) Probability of cures (b) Probability of causes
   (c) Probability of cases (d) Probability of posses

10. Two or more events are considered as ________ event. [ ]
    (a) Independent (b) Dependent
    (c) Republic (d) Transformational

II Fill in the Blanks

1. Statistical probability is also known as __________ probability.
2. Probability helps in taking effective decisions under __________ conditions.
3. The result of a random experiment is usually referred as an __________.
4. When the joint occurrence of two or more events is considered then it is known as __________.
5. Axiomatic approach was given by __________.
6. Subjective approach was initiated by __________.
7. The major theorems of probability are __________ and __________ theorems.
8. __________ probability is also known as 'single probability'.
9. The probability of certain event is __________.
10. A set is usually denoted by capital letters with or without __________.

KEY

I. Multiple Choice

1. (c)
2. (a)
3. (a)
4. (c)
5. (a)
6. (d)
7. (b)
8. (c)
9. (b)
10. (a)

II. Fill in the Blanks

1. Relative frequency probability
2. Uncertain
3. Out come
4. Compound event
5. A.N. Kolmogorov
6. Frank Ramsey
7. Addition, Multiplicative
8. Marginal probability
9. 1
10. Subscripts.

III Very Short Questions and Answers

Q1. Define probability.

Answer :

Probability can be defined as the chance or 'likelihood of occurrence' of an experiment or event. Probability of any event ranges from 0 to 1. The term 'probability' was coined by an Italian mathematician, 'Galileo'. He was the first man who measured probability quantitatively while dealing with the problems associated with rolling of dice in gambling.

Q2. What is Experiment?

Answer :

An experiment is also referred as random experiment. It is a process or activity which leads to a particular outcome of several possible outcomes. The outcome which is going to be derived through random experiment is not known until it's occurence (i.e., the outcome of random experiment is not predictable). But, the number of possible outcomes can be known. There may be fixed or infinite number of outcomes for a particular experiment. Outcome from random experiment may be numerical or non-numerical in nature.

Q3. What do you understand by set?

Answer :

The collection of finite or infinite number of objects with some common property is called Set. The objects belonging to the set are called Members or Elements of the set.

Q4. Write a note on Baye's theorem.

Answer :

Baye's theorem got its name from the British mathematician, 'Thomas Bayes' in 1763. Baye's theorem deals with revising of priori probability by making use of new information for calculating posterior probabilities. It is also known as 'Rule for the Inverse Probability'.

Q5. What are Complementary Events?

Answer :

Two events are said to be complementary events if they are mutually exclusive and collectively exhaustive.

For example, when a coin is tossed, getting a head or a tail are mutually exclusive and collectively exhaustive. Hence, if we get tail then head is considered as its complementary event.

UNIT 5 THEORETICAL DISTRIBUTIONS

SYLLABUS

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).

LEARNING OBJECTIVES

- ✓ The Concept of Binomial Distribution with its Importance, Properties, Applications and Assumptions.
- ✓ Conditions and Constants under Binomial Distribution.
- ✓ Fitting of Binomial Distribution.
- ✓ The Concept of Poisson Distribution with its Importance, Conditions and Constants.
- ✓ Fitting a Poisson Distribution.
- ✓ The Concept of Normal Distribution with its Importance and Characteristics.
- ✓ The Concept of Central Limit Theorem.
- ✓ Fitting a Normal Distribution.

INTRODUCTION

Binomial distribution is a discrete probability distribution developed by a Swiss mathematician, 'James Bernoulli' in 1700. Thus, it is also known as Bernoulli distribution. Binomial distribution is used for finite or limited number of trails 'n'. It produces successes and failures based on two parameters 'n' and 'p'.

Poisson distribution is a discrete probability distribution for countably infinite trials. Poisson distribution is named after French mathematician, 'SIMEON DENIS POISSON' in 1837. It is used when the probability of success of any individual event is very small. 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., $p \rightarrow o$).

Normal distribution was first discovered by 'Abraham Demoivre' in 1733 as a limiting case of binomial distribution. It was later developed by LAPLACE and GAUSS. Normal distribution is also known as 'Gaussian distribution' as the credit goes to German mathematician 'karl friedrich gauss'. In normal distribution, the probability of occurrence of values of random variables are calculated within a range or interval whereas the probability of a particular value cannot be calculated and is always assumed to be zero. The probability distribution used mostly for dealing with quantities whose magnitude is continuously variable is normal distribution. Hence, it is called as 'Normal probability distribution'.

PART-A
SHORT QUESTIONS AND ANSWERS

Q1. Write the Mean of Binomial Distribution.

Answer :

The mean of binomial distribution is denoted by 'μ' or '$E(X)$'. It is the expected number of successes in 'n' number of trials.

$\therefore$ The mean of binomial distribution,

$$\boxed{\mu = np}$$

Where,

'n' = Number of trials

'p' = Probability of success in a single trial.

The two parameters of binomial distribution are 'n' and 'p'.

Q2. Bring out the differences between Binomial and Poisson distribution.

Answer :

Differences between Binomial and Poisson distribution are as follows,

S.No.	Basis	Binomial Distribution	Poisson Distribution
1.	Uses	It is used for finite or limited number of trails 'n'.	It is used where 'n' $\to \infty$ and probability of success $P \to 0$.
2.	Parameters	It has two parameters 'n' and 'p'.	It has a single parameter 'λ' (where $\lambda = np$).
3.	Mean and Variance	In this, mean $\mu = np$ and variance $\sigma^2 = npq$.	In this, mean and variance are equal to $\boxed{\lambda = \sigma^2}$ or $\boxed{\sigma = \sqrt{\lambda}}$.
4.	Success and Failures	It produces successes and failures.	It produces successes commonly referred as 'occurrences'.

Q3. Write about the relation between normal and binomial distribution.

Answer :

Binomial distribution can be closely approximated to normal distribution under certain conditions which are as follows,

1. When the number of trials 'n' is very large, $n \to \infty$.
2. Either 'p' or 'q' is too close to zero.

Thus, the standardized random variable is given by,

$$\boxed{z = \frac{x - np}{\sqrt{npq}}}$$

'z' will follow normal distribution with mean 'zero' and variance 'one'.

Q4. Comment on the following:

For a Binomial Distribution Mean = 7 and Variance = 11.

Answer : (Jan.-21, Q8 (OU) | May/June-19, Q8 (OU))

For a Binomial Distribution,

Mean = $np = 7$ ......(1)

Variance = $npq = 11$ ......(2)

Dividing (2) by (1)

$$\frac{npq}{np} = \frac{11}{7}$$

$q = 1.57$ (or) 1.6

The value of q (probability) should not be more than 1. It should lie between 0 and 1. Therefore, it can be concluded that, the given data is inconsistent.

Q5. Properties of Normal Distribution.

May/June-18, Q8 (OU)

OR

Write any five properties of Normal Distribution.

Answer : July/Aug.-21, Q1(e) (OU)

Following are the characteristics/features/ properties of normal distribution,

1. The normal curve is 'bell-shaped' and symmetrical about the mean (skewness = 0). If the curve is folded along with its central vertical axis then the curves are meant to coincide with the either side of the axis.

2. The height of the normal curve is maximum at its mean. Hence, the mean and mode coincide. Thus, in normal distribution, mean, mode and median are equal.

3. The height of the curve is maximum at its mean but reduces as it goes towards either of the direction but never touches the base. Hence, the curve is known as ASYMPTOTIC. The range is unlimited or infinite in both the directions.

4. The normal curve has only one mode as there is only one maximum point which is known as 'unimodal'.

5. The points of inflexion i.e., the points where the change in curvature occurs are $\bar{x} \pm \sigma$ (or) $\mu \pm \sigma$.

Q6. The Mean of a Binomial Distribution is 4 and its Standard Deviation is $\sqrt{3}$. What are the values of n, p and q with usual notation?

Answer : Sept./Oct.-21, Q5 (OU)

Given, Mean = 4, where Mean = np; np = 4

Standard Deviation (SD) = $\sqrt{3}$,

Where, SD = $\sqrt{npq}$

Variance $(\sigma^2) = (\sqrt{3})^2 = 3$

$$npq = 3$$
$$4q = 3$$
$$q = \frac{3}{4}$$

We know that, $p + q = 1$

$$p + \frac{3}{4} = 1$$
$$p = 1 - \frac{3}{4}$$
$$p = \frac{1}{4}$$

$$np = 4, n\left(\frac{1}{4}\right) = 4, n = 4 \times 4 = 16$$

$$\therefore p = \frac{1}{4}, q = \frac{3}{4}, n = 16.$$

Q7. 6 coins are tossed at the same time find the probability that 4 heads are occurred.

Answer : May/June-19, Q5 (MGU)

Given that,

Total number of coins tossed at a time, n = 6

Probability of getting head P(H) = $\frac{1}{2}$

Probability of getting tail P(T) = $\frac{1}{2}$

Probability of getting 4 heads P(H = 4),

$$= {}^nC_r\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{n-r}$$
$$= {}^6C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{6-4}$$
$$= \frac{6!}{2!\times 4!}\times\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^2$$
$$= \frac{6\times5\times4!}{2!\times4!}\times\left[\frac{1}{2}\right]^4\left[\frac{1}{2}\right]^2$$
$$= 15\times\left(\frac{1}{2}\right)^6$$
$$= 15\times\left[\frac{1}{64}\right]$$
$$= \frac{15}{64}$$

$\therefore$ The probability of getting 4 heads P(H = 4) = $\frac{15}{64}$.

PART-B

ESSAY QUESTIONS AND ANSWERS

5.1 PROBABILITY OR THEORETICAL DISTRIBUTION

Q8. Write a short note on probability or theoretical distribution and state its uses.

Answer :

Probability distribution is a set of probabilities of all the possible outcomes of a random experiment. In other words the desire outcome derived from related probabilities is called as "Probability Distributions". It is based on the theoretical considerations, subjective assessment or on experience as such it is also called as Theoretical Distribution. Probability distribution functions are usually represented as p(x), f(x), h(x) etc.

Types of Theoretical Probability Distributions

The types of theoretical probability distributions are as follows,

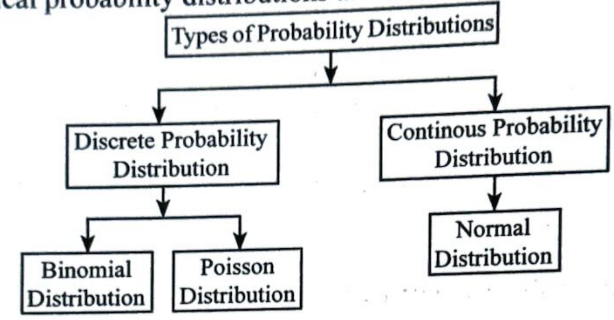


Figure: Types of Probability Distributions

1. Discrete Probability Distribution

Discrete probability deals with limited or finite number of outcomes. It is very easy to calculate probability through this method as such it is commonly used for practical solutions.

(i) Binomial Distribution

Binomial distribution is a discrete probability distribution developed by a Swiss mathematician, 'James Bernoulli' in 1700. Thus, it is also known as Bernoulli distribution. Binomial distribution is used for finite or limited number of trials 'n'. It produces successes and failures based on two parameters 'n' and 'p'.

(ii) Poisson Distribution

Poisson distribution is a discrete probability distribution for countably infinite trials. Poisson distribution is named after French mathematician, 'SIMEON DENIS POISSON' in 1837. Poisson Distribution is used when the probability of success of any individual event is every small. The average or mean of poisson distribution is given by λ.

2. Continous Probability Distribution

Continous probability distribution deals with unlimited/infinite number of outcomes. It is mostly used for realistic description of outcome. Normal distribution is the important type or method of continous probability distribution.

❖ **Normal Distribution**

Normal distribution was first discovered by 'Abraham Demoivre' in 1733 as a limiting case of binomial distribution. It was later developed by LAPLACE and GAUSS. Normal distribution is also known as 'Gaussian distribution' as the credit goes to German mathematician 'karl friedrich gauss'. In normal distribution, the probability of occurrence of values of random variables are calculated within a range or interval whereas the probability of a particular value cannot be calculated and is always assumed to be zero. The probability distributions used mostly for dealing with quantities whose magnitude is continuously variable as normal distribution. Thus, it is called as "Normal probability distribution". It plays a prominent role in statistical theory and practice.

Uses of Probability/Theoretical Distribution

The following points highlights the uses or importance of probability/theoretical distribution,

1. It is used to make decisions about the future variables.
2. It is used to forecast and predict the business.
3. It is used in analyzing the actual and expected frequencies to know about the differences between them is real or imaginary.
4. It is used by the businessmen, bankers, government, industrialists, insurance companies, social workers, transporters etc as it helps to solve the various problems.
5. It is useful for examining risk and uncertainty of the business.
6. It is used when actual distributions are expensive and unable to acquire.

5.1.1 Binomial Distribution – Importance, Properties, Applications and Assumptions

Q9. What is Binomial distribution? State its importance, applications and assumptions.

Answer :

Binomial Distribution

Binomial distribution is a discrete probability distribution developed by a Swiss Mathematician, 'James Bernoulli' in 1700. Thus, it is also known as Bernoulli distribution. It is used for finite or limited number of trials 'n'. It produces successes and failures based on two parameters 'n' and 'p'.

Binomial distribution satisfies two essential properties of probability distribution which are explained as follows,

(i) $f(x) \geq 0$

Binomial distribution fulfills this requirement as 'n' and 'p' both are positive.

Therefore, ${}^nC_r\, p^r\, q^{n-r}$ are all positive.

So $\boxed{f(x) \geq 0}$

(ii) $\Sigma f(x) = 1$

Binomial expansion of $(p+q)^n$ helps in fulfilling this requirement.

$$\text{As } \Sigma f(x) = \Sigma {}^nC_r . p^r q^{n-r} \quad [\because p = 1 - q]$$

$$= (p+q)^r \Rightarrow (1 - q + q)^r$$

$$= (1)^r = 1 \Rightarrow \boxed{\Sigma f(x) = 1}$$

Importance of Binomial Distribution

The following points highlights the importance of binomial distribution,

1. It is an extensively used probability distribution of a discrete random variable.
2. It plays a vital role in the functions of quality control and quality assurance.
3. It is used in manufacturing units for defective analysis.
4. It is also used in service organizations like banks and insurance companies.
5. It characterizes the outcomes of each trials in the process on one of two types of possible outcomes.
6. The possibility of outcome of any trials does not change and remains independent compared to previous trials.

Applications of Binomial Distribution

Binomial distribution is applicable in case of repeated trials such as,

1. Number of applications received for a junior assistant post during a particular period of time.
2. Number of births taking place in a hospital.
3. Number of candidates appearing for the screening test conducted by a company.

Thus, all the trials are statistically independent and each trial has two outcomes namely, success and failure.

Assumptions of Binomial Distribution

Binomial distribution is assumed under the following conditions,

1. The number of trials 'n' is fixed and finite.
2. The trials are independent of each other.
3. The probability of success 'p' is constant for each trial and also the probability of failure 'q' is constant for each trial. Always $\boxed{p+q=1}$
4. The trial has only two possible outcomes such as success and failure.

Q10. Explain about the properties, mean and variance of binomial distribution.

Answer :

Properties of Binomial Distribution

The properties of Binomial distribution are as follows,

1. **Distribution of Probabilities**

   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. **Identical Conditions**

   The process is performed under identical conditions for 'n' number of times.

3. **Independent**

   Each trial is independent of other trials. It means the outcome of a particular trial does not affect the outcome of another trial.

4. **Observation**

   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. **Symmetrical**

   Binomial distribution is symmetrical when $p = 0.5$ [figure (i)] and it is skewed if $p \neq 0.5$, where 'n' can be any value.

   When $p > 0.5$ [figure (iii)], it is skewed to the right $\rightarrow$ negatively skewed.

   When $p < 0.5$ [figure (ii)], it is skewed to the left $\rightarrow$ positively skewed.

   Hence, binomial distribution is 'Asymmetrical'

   When $p > 0.5$ and $p < 0.5$

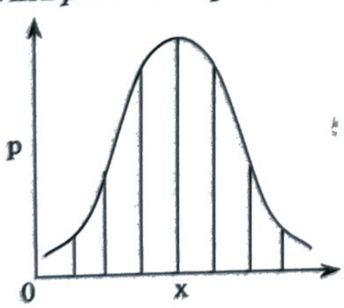


Figure (i)

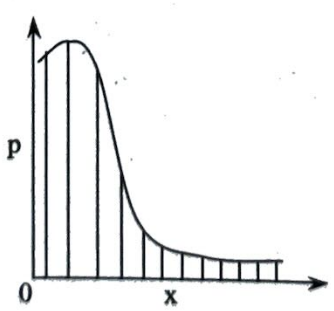


Figure (ii)

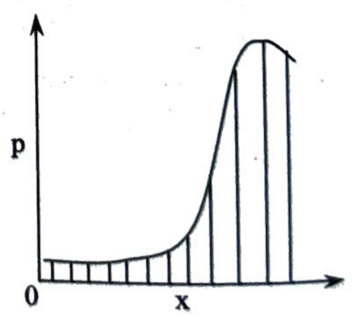


Figure (iii)

6. **Standardization**

   If 'n' is large and if neither 'p' nor 'q' is nearly zero, in such cases the binomial distribution is modified to normal distribution by standardizing the variable, $\boxed{Z = \frac{X - np}{\sqrt{npq}}}$

Mean of Binomial Distribution

The mean of binomial distribution is denoted by 'μ' or '$E(X)$'. It is the expected number of successes in 'n' number of trials.

∴ The mean of binomial distribution,

$$\mu = np$$

Where,

'n' = Number of trials

'p' = Probability of success in a single trial.

The two parameters of binomial distribution are 'n' and 'p'.

Variance of Binomial Distribution

The variance of binomial distribution is denoted by 'σ^2'. It is the square of the standard deviation.

Thus, the variance of binomial distribution is given by,

$$\sigma^2 = npq$$

Where,

n = Number of trials

p = Probability of success in a single trial

q = Probability of failure in a single trial.

Standard deviation is given by,

$$\sigma = \sqrt{npq}$$

$$\Rightarrow \text{Standard deviation} = \sqrt{\text{Variance}}$$

$$\Rightarrow \text{Variance, } \sigma^2 = \mu . q \quad [\because \mu = np]$$

Q11. What do you mean by Binomial Distribution? Is there any fallacy in the statement the mean of a Binomial Distribution is 20 and its Standard Deviation is 7.

Answer : Sept/Oct.-21, Q18 (OU)

Binomial Distribution

For answer refer Unit-V, Page No. 131, Q.No. 9, Topic: Binomial Distribution.

Statement

Given Mean = np = 20

Standard deviation (σ) = $\sqrt{npq}$ = 7

σ^2 = npq = 49

20q = 49

q = $\frac{49}{20}$

q = 2.45

∴ The value of q should lies between 0 and 1. But, the value exceeds the given limit. So, the given statement is wrong.

5.1.2 Conditions and Constants Under Binomial Distribution

Q12. Write about the conditions and constants under which binomial distribution is used.

Answer :

Conditions Under which Binomial Distribution is Used

The following are the conditions which must be fulfilled for using the binomial distribution,

1. The trials has two mutually exclusive outcomes and collective exhaustive outcomes which are referred as "Success" and "Failure".
2. The outcomes of the trial 'success' is denoted by 'P' and failure denoted by 'q'.
3. The experiment is repeated under the same conditions for a fixed and finite number of times.
4. The observation of random variable in an random experiment is known as trial.
5. The possibility of outcome of any trial does not change and remains constant compared to previous trials.
6. The trials are independent i.e., outcomes of one trial has no effect on the outcome of other.

Constants of Binomial Distribution

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 trial

q = Probability of failure on a single trial

n = Number of Bernoulli trials

Mean : $\mu = np$, Variance : $\sigma^2 = npq$

The following are the constants of binomial distribution,

(i) The binomial distribution variance is less than mean.

(ii) The binomial distribution with parameters n and p variance cannot exceed $n/4$.

(iii) The 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_3)	$= \dfrac{q-p}{\sqrt{npq}}$
Kurtosis (μ_4)	$= \dfrac{1-6pq}{npq}$

5.1.3 Fitting of Binomial Distribution

Q13. Briefly describe about fitting a binomial distribution along with an illustration.

Answer :

Fitting a Binomial Distribution

The following are the steps to be considered while fitting a binomial distribution,

Step-1

Calculate the values 'p' and 'q' if one is known other can be obtained as $p + q = 1$

Step-2

Expand binomial $(p + q)^n$

Where,

n = One less than the number of terms obtained after expansion.

For example, if 'n' value is 3, then number of terms obtained after expansion is 4.

Step-3

Multiply each term of the expanded binomial by 'N' which is sum of all the frequencies to obtain the expected individual frequencies.

Illustration

For answer refer Unit-V, Page No. 134, Q.No. 14.

PROBLEMS ON BINOMIAL DISTRIBUTION

Q14. Fit a binomial distribution,

X	0	1	2	3	4	5	6	7
Y	7	6	19	35	30	23	7	1

Solution :

Steps in the Fitting of Binomial Distribution

Step-1

Calculate the values p and q

Mean of binomial distribution = np

Mean of frequency distribution = $\dfrac{\Sigma fx}{\Sigma f}$

Here 'X' is the random variable for success and the number of Bernoulli trials here is $n = 7$ (As there are 8 terms, 'n' value is always taken as one less than the number of terms).

Mean can be calculated by the given below,

X	f	f(x)
0	7	0
1	6	6
2	19	38
3	35	105
4	30	120
5	23	115
6	7	42
7	1	7
Total	$\Sigma f = 128$	$\Sigma f(x) = 433$

From the above table, $\Sigma f(x) = 433$

$\Sigma f = 128$

$$\text{Mean} = \frac{\Sigma f(x)}{\Sigma f}$$

$$= \frac{433}{128}$$

$$= 3.38$$

$\therefore$ Mean = 3.38 ... (1)

Mean = np ... (2)

[From mean of binomial distribution]

From equations (1) and (2) we get,

$\therefore$ $np = 3.38$

[$\because$ $n = 7$ (from given problem)]

$\therefore$ $7p = 3.38$

$$p = \frac{3.38}{7}$$

$= 0.48$

∴ The probability of success, $p = 0.48$

Probability of failure, $q = 1 - 0.48$

$[\because p + q = 1]$

$= 0.52$

∴ $p = 0.48$ and $q = 0.52$

Step-2

Expand Binomial $(p + q)^n$

The expected binomial probabilities are calculated by using the formula of Bernoulli distribution.

This is given by,

$$p(r) = {}^nC_r p^r q^{n-r}$$

$n = 7$ (∵ from table) $p = 0.48$

$r = 0, 1, 2, 3, 4, 5, 6, 7$ (given) $q = 0.52$

$p(r) = {}^nC_r p^r q^{n-r}$

$p(0) = {}^7C_0(0.48)^0(0.52)^7 \quad [\because {}^nC_0 = 1]$

$= 1 \times 1 \times 0.01 = 0.01 \quad [x^0 = 1]$

∴ $p(0) = 0.01$

$p(1) = {}^7C_1(0.48)^1(0.52)^6 \quad [\because {}^nC_1 = n]$

$= 7 \times 0.48 \times 0.019$

∴ $p(1) = 0.064 \left[{}^nC_r = \frac{n!}{r!(n-r)!}\right]$

$p(2) = {}^7C_2 (0.48)^2 (0.52)^5$

$= \frac{7\times6}{2} \times 0.23 \times 0.038$

$= 21 \times 0.23 \times 0.038$

∴ $p(2) = 0.18$

$p(3) = {}^7C_3(0.48)^3(0.52)^4$

$= \frac{7\times6\times5}{3\times2} \times 0.11 \times 0.073$

$= 35 \times 0.11 \times 0.073$

∴ $p(3) = 0.28$

$p(4) = {}^7C_4(0.48)^4(0.52)^3 \quad [{}^7C_3 = {}^7C_4 = 35]$

$= 35 \times 0.053 \times 0.14$

∴ $p(4) = 0.26$

$p(5) = {}^7C_5(0.48)^5(0.52)^2 \quad [{}^7C_5 = {}^7C_2 = 21]$

$= 21 \times 0.026 \times 0.27$

∴ $p(5) = 0.15$

$p(6) = {}^7C_6(0.48)^6(0.52)^1 \quad [{}^7C_6 = {}^7C_1 = 7]$

$= 7 \times 0.012 \times 0.52$

∴ $p(6) = 0.04$

$p(7) = {}^7C_7(0.48)^7(0.52)^0 \quad [\because {}^nC_n = 1]$

$= 1 \times 0.006 \times 1 \quad [x^0 = 1]$

∴ $p(7) = 0.006$

Step-3

Multiply each term with total frequency (N) to obtain expected frequencies.

The expected frequencies of the distribution are given by,

$$f(r) = N.p(r)$$

Filling of Binomial Distribution

r	$p(r) = {}^nC_r p^r q^{n-r}$	$f(r) = N.p(r) = 128.p(r)$
0	$p(0) = {}^7C_0(0.48)^0(0.52)^7 = 0.01$	$f(0) = 128 \times 0.01 = 1.28$
1	$p(1) = {}^7C_1(0.48)^1(0.52)^6 = 0.064$	$f(1) = 128 \times 0.064 = 8.192$
2	$p(2) = {}^7C_2(0.48)^2(0.52)^5 = 0.18$	$f(2) = 128 \times 0.18 = 23.04$
3	$p(3) = {}^7C_3(0.48)^3(0.52)^4 = 0.28$	$f(3) = 128 \times 0.28 = 35.84$
4	$p(4) = {}^7C_4(0.48)^4(0.52)^3 = 0.26$	$f(4) = 128 \times 0.26 = 33.28$
5	$p(5) = {}^7C_5(0.48)^5(0.52)^2 = 0.15$	$f(5) = 128 \times 0.15 = 19.2$
6	$p(6) = {}^7C_6(0.48)^6(0.52)^1 = 0.04$	$f(6) = 128 \times 0.04 = 5.12$
7	$p(7) = {}^7C_7(0.48)^7(0.52)^0 = 0.006$	$f(7) = 128 \times 0.006 = 0.768$

Note

1. Always the sum of all the probabilities is equal to 1.
2. Always the sum of all the calculated expected frequencies is equal to the sum of frequencies ($N = \Sigma f$) mentioned in the problem.

Q15. 8 Coins are tossed at a time, 256 times. Find the expected frequencies of successes (Getting a Head) and tabulate the results obtained.

Solution : Jan.-21, Q17 (OU)

Given that,

$n = 8$

$N = 256$

The Probability of getting a head $(p) = \frac{1}{2}$

The Probability of getting a tail $(q) = \frac{1}{2}$

The Probability of Success r times in n trials is given by ${}^nC_r q^{n-r} p^r$

$\therefore P(r) = {}^nC_r\, q^{n-r}\, p^r$

$= {}^8C_r \left(\frac{1}{2}\right)^{8-r} \left(\frac{1}{2}\right)^r$

$= {}^8C_r \left(\frac{1}{2}\right)^r$

The frequencies of 0,1,2,3,...,8 Successes 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
	Total	**256**

Working Notes

Sample Calculation of N × P (r)

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

${}^8C_3 = \frac{8!}{3!(8-3)!}$

${}^8C_3 = \frac{8!}{3!(5)!}$

${}^8C_3 = \frac{{}^4\cancel{8}\times 7\times {}^2\cancel{6}\times\cancel{5\times 4\times 3\times 2\times 1}}{\cancel{3}_1\times\cancel{2}_1\times 1\,\cancel{(5\times 4\times 3\times 2\times 1)}}$

$= \frac{4\times 7\times 2}{1\times 1\times 1} = \frac{4\times 7\times 2}{1} = 56$

$= 256\left(\frac{1}{256}\times 8c_3\right)$

$= 256\left(\frac{1}{256}\times 56\right)$

$= 256\,(0.21875)$

$= 56$

(ii) ${}^8C_6 = \frac{8!}{6!\,(8-6)!}$

$= \frac{8\times 7\times\cancel{6\times 5\times 4\times 3\times 2\times 1}}{\cancel{6\times 5\times 4\times 3\times 2\times 1}\,(2)!}$

$= \frac{\cancel{8}^4\times 7}{{}_1\cancel{2}\times 1} = 28$

$= 256\left(\frac{1}{256}\times 28\right) = 28$

Q16. 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.

Solution : May/June-19, Q13(a) (OU)

According to Binomial Probability of law, Probability of r heads is given by,

$P(r) = P\,(X = r) = {}^nC_r p^r q^{n-r}$

Where, ${}^nC_r = \frac{n!}{r!\left(n-r\right)!}$

P = Probability of obtaining head = $\frac{1}{2}$

q = Probability of obtaining tail = $\frac{1}{2}$

$n = 10$

(i) Probability of Obtaining Exactly 6 Heads

$n = 10, r = 6, p = \frac{1}{2}, q = \frac{1}{2}$

$P(\text{6 heads}) = {}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^{10-6}$

$= {}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4$

${}^{10}C_6 = \frac{10!}{6!\left(10-6\right)!} = \frac{10!}{6!\left(4!\right)}$

$= \frac{10\times 9\times 8\times 7\times 6!}{6!\left(4\times 3\times 2\times 1\right)}$

$= 210$

$\therefore P(\text{6 heads}) = 210\left(\frac{1}{64}\right)\left(\frac{1}{16}\right)$

$= \frac{210}{1,024}$

$= 0.205$

$\therefore$ Probability of obtaining exactly 6 heads is 0.205.

(ii) Probability of Obtaining Atleast Eight Heads

$P(X \geq 8) = P(8\text{ heads}) + P(9\text{ heads}) + P(10\text{ heads})$

$$P(8\text{ heads}) = {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^{10-8}$$
$$= {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2$$
$$= \frac{10!}{8!(10-8!)} \left(\frac{1}{256}\right)\left(\frac{1}{4}\right)$$
$$= \frac{10\times 9\times 8!}{8!(2!)} \left(\frac{1}{256}\right)\left(\frac{1}{4}\right)$$
$$= 45 \left(\frac{1}{256}\right)\left(\frac{1}{4}\right)$$
$$= \frac{45}{1024}$$
$$= 0.044$$

$$P(9\text{ heads}) = {}^{10}C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^{10-9}$$
$$= \frac{10!}{9!(10-9!)} \left(\frac{1}{512}\right)\left(\frac{1}{2}\right)$$
$$= \frac{10\times 9!}{9!(1)} \left(\frac{1}{1024}\right)$$
$$= \frac{10}{1024}$$
$$= 0.01$$

$$P(10\text{ heads}) = {}^{10}C_{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^{10-10}$$
$$= 1\left(\frac{1}{1024}\right)(1)$$
$$= 0.001$$

Now, $P(X \geq 8) = 0.044 + 0.01 + 0.001$

$= 0.055$

∴ Probability of obtaining atleast 8 heads is 0.055.

(iii) Probability of Obtaining No Heads

$$P(0\text{ heads}) = {}^{10}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10-0}$$
$$= 1\,(1)\left(\frac{1}{1024}\right)$$
$$= 0.001$$

∴ Probability of obtaining no heads is 0.001

(iv) Probability of Obtaining Atleast One Head

$P(\text{atleast one head}) = 1 - P(\text{no heads})$

[From (iii), P(no heads)] $= 0.001] = 1 - 0.001$

$= 0.99$

∴ Probability of obtaining atleast one head is 0.001

(v) Probability of Obtaining Not More Than 3 Heads

$P(X \leq 3) = P(0\text{ heads}) + P(1\text{ heads}) + P(2\text{ heads}) + P(3\text{ heads})$

$P(0\text{ heads}) = 0.001$

$$P(1\text{ head}) = {}^{10}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{10-1}$$
$$= 10\left(\frac{1}{2}\right)\left(\frac{1}{512}\right) = \frac{10}{1024}$$
$$= 0.01$$

$$P(2\text{ heads}) = {}^{10}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{10-2}$$
$$= \frac{10!}{8!\times 2!}\left(\frac{1}{4}\right)\left(\frac{1}{256}\right)$$
$$= \frac{10\times 9\times 8!}{8!\times 2\times 1}\left(\frac{1}{1024}\right)$$
$$= 45\left(\frac{1}{1024}\right)$$
$$= 0.044$$

$$P(3\text{ heads}) = {}^{10}C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{10-3}$$
$$= \frac{10!}{7!\times 3!}\left(\frac{1}{8}\right)\left(\frac{1}{128}\right)$$
$$= \frac{10\times 9\times 8\times 7!}{7!\times 3\times 2\times 1}\left(\frac{1}{1024}\right)$$
$$= \left(\frac{120}{1024}\right)$$
$$= 0.12$$

$P(X \leq 3) = 0.001 + 0.01 + 0.044 + 0.12$

$= 0.175$

∴ Probability of obtaining not more than 3 heads is 0.175

(vi) Probability of Obtaining Atleast Four Heads

$P(X \geq 4) = 1 - P(X \leq 3)$

[From (v) $P(X \leq 3) = 0.175$]

Now, $P(X \geq 4) = 1 - 0.175$

$= 0.825$

∴ Probability of obtaining atleast 4 heads is 0.825.

Q17. In a city half of the population are Rice consumers to find this truth 100 supervisors are appointed. Every supervisor has been examined 10 members, what is the probability to three or less than three people are rice consumers reports supervisors number.

May/June-19, Q10(b) (MGU)

Solution :

Note: In the given question, there is incomplete information due to which we have revised the question in order to solve it. The revised question (original question of text book) is as follows,

Suppose that half the population of a town are consumers of rice. 100 investigators are appointed to find out its truth. Each investigator interviews 10 individuals. How many investigators do you expect to report that three or less of the people interviewed are consumers of rice? July/Aug.-21, Q10 (KU)

Let 'x' be the probability of persons out of 10 persons who are rice consumers.

From binomial distribution,

$$P(X=x) = {}^nC_x\, p^x q^{n-x}$$

Given that,

$$p = \frac{1}{2}, q = 1-p = \frac{1}{2}, n = 10$$

$$P(X=x) = {}^{10}C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{10-x}$$

∴ The probability that there are 3 people or less who are rice consumers.

$$p(X\leq 3) = p(x=0)+p(x=1)+p(x=2)+p(x=3)$$

$$p(X=0) = {}^{10}C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{10-0} = 1(1)(1/2)^{10} = \frac{1}{2^{10}}$$

$$p(X=1) = {}^{10}C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{10-1} = 10(1/2)(1/2)^9 = \frac{10}{2^{10}}$$

$$p(X=2) = {}^{10}C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{10-2} = 45(1/2)^2(1/2)^8 = \frac{45}{2^{10}}$$

$$p(X=3) = {}^{10}C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{10-3} = 120(1/2)^3(1/2)^7 = \frac{120}{2^{10}}$$

$$p(X\leq 3) = \frac{1}{2^{10}}+\frac{10}{2^{10}}+\frac{45}{2^{10}}+\frac{120}{2^{10}}$$

$$= \frac{1+10+45+120}{2^{10}} = \frac{176}{1024} = \frac{11}{64}$$

$$= 0.171875$$

$$\simeq 0.172$$

∴ The expected number of supervisors who would report that there are '3' or less people were rice consumers.

$$= 100\times 0.172 = 17.2$$

$$\simeq 17.$$

5.2 POISSON DISTRIBUTION – IMPORTANCE, PROPERTIES AND APPLICATIONS

Q18. Define poisson distribution and state its importance.

Answer :

Poisson Distribution

Poisson distribution is a discrete probability distribution for countably infinite trials. Poisson distribution is named after French mathematician, 'SIMEON DENIS POISSON' in 1837. It is used when the probability of success of any individual event is very small.

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., $p\to o$)

Always the sum of infinite probabilites in poisson distribution is 1 i.e.,

$$\boxed{P(0)+P(1)+P(2)+...+P(\infty)=1}$$

Importance of Poisson Distribution

Following points highlight the importance of Poisson Distribution,

1. It helps to describe the complete randomness and independence of events.
2. It helps to make unlimited number of trials.
3. It helps to find out the probability of events occurring in a fixed interval of time.
4. It is more effective and helpful compare to binomial distribution because it requires only Mean (M) whereas binomial distribution requires n and p.
5. It helps to find out the optimal size of a unit.
6. It helps in understanding the problems and finding their solutions.
7. It helps events with low probabilities of occurrence within some definite time or space.
8. It helps businessmen to make forecasts for number of customers or sales on certain days or seasons.

Q19. Explain about the properties and applications of poisson distribution.

OR

Explain the features of Poisson distribution.

May/June-19, Q10(a) (MGU)

(Refer Only Topic: Properties/Features of Poisson Distribution)

Answer :

Properties/Features of Poisson Distribution

Following are the properties/features of Poisson distribution,

1. Independent

The occurrence of the event is independent i.e., the occurrence of an event in a time interval has no effect on the occurrence of the second event in the same or any other interval.

2. Infinite Number

Theoretically, an infinite number of occurrences of the event must be possible in the interval.

3. Single Occurrence

The probability of single occurrence of the event in a given interval is directly proportional to the length of the interval.

4. Small Portion

In any extremely small portion of the interval, the probability of two or more occurrences of the event is negligible. Like binomial distribution, poisson distribution also satisfies the two essential properties i.e.,

(i) $f(x) \geq 0$ and

(ii) $\Sigma f(x) = 1$.

5. Discrete Probability

It is a discrete probability distribution.

6. Skewed

It is positively skewed to right.

Applications of Poisson Distribution

Poisson distribution is mostly applied in business, management science and operations research. Some of the examples of applications which are observed in our daily life where poisson distribution is used are as follows,

(i) Number of calls received at a call centre.

(ii) Number of printing mistakes occurred on the pages in a book.

(iii) Number of trains arrived at railway station.

(iv) Number of persons joining a queue at a bank.

(v) Number of bacteria in a given media.

(vi) Number of typing mistakes detected by proofing department.

(vii) Number of particles emitted by a radio active substance.

(viii) Number of defective items identified in box containing very large number of items.

(ix) Number of customers served at a telephone department.

(x) Number of deaths in a village by an unknown disease.

5.2.1 Conditions and Constants Under Poisson Distribution

Q20. Write about the conditions and constants under which Poisson Distribution is used.

Answer :

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. $np = \lambda$ is a finite constant (positive real number)

Thus, $p = \frac{\lambda}{n}$ and $q = \left(1 - \frac{\lambda}{n}\right)$

Probability of 'x' success in a series of 'n' independent trials is,

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

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,

$$P(X) = \frac{e^{-\lambda}.\lambda^x}{x!}$$

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'

$$\lambda = np$$

'λ' is the single parameter of Poisson distribution. As 'λ' increases, the distribution shifts to right.

Hence, it is called as 'Probability distribution of rare events' or "Law of improbable events".

It 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.

It is a distribution of rare events. Here the finite number of trials i.e., the value of 'n' is not mentioned.

'n' tends to ∞ in this case.

Under Poisson distribution,

$$\text{Mean} = \text{Variance} = \lambda$$

5.2.2 Fitting of Poisson Distribution

Q21. Explain the steps of fitting a Poisson Distribution.

Answer :

The following are the steps for fitting a Poisson distribution,

Step-1

Calculate the values of mean (λ) and probability of zero occurrence.

Mean in poisson distribution is calculated as,

$$\lambda = nxp$$

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 (N) to obtain expected frequency values.

PROBLEMS ON POISSON DISTRIBUTION

Q22. Fit a poisson distribution to the following data.

x:	0	1	2	3	4
f:	123	59	14	3	1

(e^{-m} = 0.6065).

Solution : Jan.-21, Q18 (OU)

Step-1

Calculate the values of 'λ' and Probability of zero Occurrence.

X	f	fx
0	123	0
1	59	59
2	14	28
3	3	9
4	1	4
	$\Sigma f = 200$	$\Sigma fx = 100$

Mean of poisson distribution is given

$\lambda = \frac{\Sigma fx}{\Sigma f}$

$\lambda = \frac{100}{200} = 0.50$

$\therefore \lambda = 0.50$

Step-2

Calculate all the probabilities by using recurrence relation,

$P(0) = 0.6065$ ($\therefore$ Given e^{-m} (or) $e^{-\lambda}$ 0.6065)

$P(1) = \frac{P(0)\times\lambda}{1} \Rightarrow 0.6065\times0.5 = 0.3032$

$P(2) = \frac{P(1)\times\lambda}{2} = \frac{0.3032\times0.5}{2} = 0.0758$

$P(3) = \frac{P(2)\times\lambda}{3} = \frac{0.0758\times0.5}{3} = 0.013$

$P(4) = \frac{P(3)\times\lambda}{4} = \frac{0.013\times0.5}{4} = 0.002$

Step-3

Multiply each term of probability with total frequency (Σf) to obtain the values of expected frequencies.

Computation of Expected Frequencies

X	P(x)	f(x) = N.P (x) = 200 P(x)
0	$P(0) = 0.6065$	$f(0) = 200 \times 0.6065 = 121.3 \cong 121$
1	$P(1) = 0.3032$	$f(1) = 200 \times 0.3032 = 60.64 \cong 61$
2	$P(2) = 0.08$	$f(2) = 200 \times 0.0758 = 15.16 \cong 15$
3	$P(3) = 0.013$	$f(3) = 200 \times 0.013 = 2.6 \cong 3$
4	$P(4) = 0.001$	$f(4) = 200 \times 0.001\ 0.2 \cong 0$
	Total	**200**

$\therefore$ The theoretically fitted poisson distribution is as follows,

x	0	1	2	3	4
y	121	61	15	3	0

Q23. In a town 10 accidents take place in span of 50 days. Assuming that the number of accidents follows poisson distribution. Find the probability that there will be 3 or more accidents in a day.

Solution :

July/Aug.-21, Q10 (MGU)

Let 'n' be the number of accidents taken place in a span of 50 days and 'p' be the probability that an accident takes place.

Given,

$n = 10$ and $p = \frac{1}{50}$

Since, 'p' is very small, Poisson distribution is applied in this case.

Mean of Poisson distribution, $\lambda = np$

$\therefore \quad \lambda = 10 \times \frac{1}{50} = 0.2$

Probability in Poisson distribution is given by,

$$P(X)=\frac{e^{-\lambda}.\lambda^x}{x!}$$

Probability that there will be 3 or more accidents in a day is given by,

$P(X \leq 3)$

$$P(0)+P(1)+P(2)+...+P(\infty)=1$$

(Sum of infinite probabilities of Poisson distribution is equal to 1)

$P(X \geq 3) = P(3) + P(4) +...+ P(\infty)$ ... (1)

$P(3) + P(4) +...+ P(\infty) = 1 - [P(0) + P(1) + P(2)]$ ... (2)

From equations (1) and (2), we get,

$P(X \geq 3) = 1 - [P(0) + P(1) + P(2)]$

$$P(X)=\frac{e^{-\lambda}.\lambda^x}{x!}$$

$\lambda = 0.2$

$e^{-\lambda} = e^{-0.2} = 0.819$

$x = 0, 1, 2$

$\therefore$ $P(X \geq 3)$

$$=1-\left[\left(\frac{e^{-0.2}(0.2)^0}{0!}\right)+\left(\frac{e^{-0.2}(0.2)^1}{1!}\right)+\left(\frac{e^{-0.2}(0.2)^2}{2!}\right)\right]$$

$= 1 - \left[e^{-0.2}(1+0.2+0.02)\right]$ $[\because x^0 = 1, 0! = 1, 1! = 1]$

$= [1 - (0.819 \times 1.22)]$ $e^{-0.2} = 0.819$

$= (1 - 0.999) = 0.001$

$\therefore$ The probability that there will be three or more accidents in a day = 0.001 or 0.1%

5.3 NORMAL DISTRIBUTION – IMPORTANCE

Q24. What is Normal Distribution? Write the importance and applications of normal distribution.

Answer :

Normal Distribution

Normal distribution was first discovered by 'Abraham Demoivre' in 1733 as a limiting case of binomial distribution. It was later developed by LAPLACE and GAUSS. It is also known as 'Gaussian distribution' as the credit goes to German mathematician 'Karl Friedrich Gauss'.

In normal distribution, the probability of occurrence of values to the random variables are calculated within a range or interval whereas the probability of a particular value cannot be calculated and is always assumed to be zero. The probability distribution used mostly for dealing with quantities whose magnitude is continuously variable as normal distribution. Thus, it is called as 'Normal probability distribution'. It plays a prominent role in statistical theory and practice.

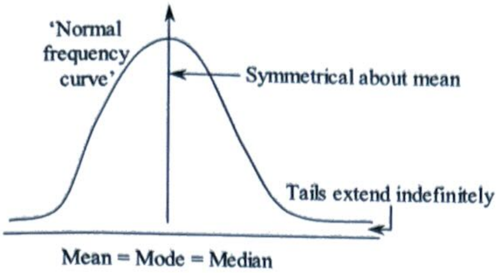
It is an approximation to binomial distribution whether p is equal to q or not. Binomial distribution tends to form a continuous curve when 'n' becomes large at least for a certain range.

Thus, the limiting frequency curve which is obtained when 'n' has large value is called as 'the normal frequency curve' or 'the normal curve'.

The two parameters of normal distribution are 'μ' (mean) and 'σ^2' (variance).

Always the value of random variable 'X' lies within certain range and has no particular value.

Thus, $-\infty < X < \infty$



Importance of Normal Distribution

Following points highlights the importance of normal distribution,

1. It is used to calculate many of the distributions like Binomial, Poisson etc., which can be approximated by normal distribution under suitable conditions.
2. It is used to fit a distribution under certain conditions.
3. It helps to calculate many of the distributions of sample statistics i.e., the distribution of sample mean, sample variance etc., tend to normality for large samples.
4. It is used for large applications in statistical quality control in industry for establishing control limits.
5. It is widely or extensively used in the entire theory of small sample tests in which it is assumed that the parent population from which the samples have been drawn follow normal distribution.
6. It plays a major role in appropriate decision making under sampling theory.
7. It helps the sampling distributions of statistics such as, students t-distribution, Snedecor's F-distribution, Fisher's Z-distribution and Chi square distribution; to adjust into normal distributions for large degrees of freedom.

Applications of Normal Distribution

Normal distribution plays a prominent role in statistical theory as well as practical area of applications. Some of the applications are as follows,

1. It is used in astronomical observations for analyzing the measurement errors.
2. It is used to determine the blood pressure of human body.
3. It is used to determine the height of an individual.

Q25. Write about,

(i) General model of normal distribution

(ii) Standard normal distribution.

Answer :

(i) General Model of Normal Distribution

Under normal distribution, the random variable 'X' is given as follows,

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

Where,

X = Values of the continuous random variable $-\infty < X < \infty$

μ = Mean of the random variable

e = Mathematical constant ($e = 2.7183$)

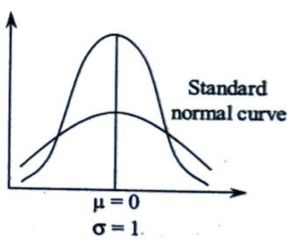
π = Mathematical constant ($\pi = 3.14$)

($2\pi = 2.5066$)

σ = Standard deviation.

Standard normal curve is a curve in which whose $\mu = 0$ and $\sigma = 1$

It is known that the normal curve has unit area, hence, $\sigma = 1$



The graph of $f(X)$ is a famous 'bell-shaped curve'.

As the curve has unit area, the total frequency (N) is equal to 1.

It is not possible to draw normal curves in all cases, hence it is necessary to standardize the normal curve which is known as 'standard normal curve'.

Thus, 'X' is converted into 'Z' known as '*standard normal variate*'.

Standard normal variate 'Z' is given by,

$$z = \frac{X-\mu}{\sigma}$$

Where,

X = Value of the observation

μ = Mean

σ = Standard deviation to F(Z).

$F(X)$ is also changed to $F(Z)$

$$F(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}}$$

(ii) Standard Normal Distribution

The distribution of a normal random variable with mean (μ) = 0 and standard deviation (σ) = 1 is called as 'standard normal distribution' and the curve is called as 'standard normal curve'.

The random variable 'x' is said to have a normal distribution with the two parameters 'μ' and 'σ' if its probability function is given by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

These two parameters describe the complete situation. The probability function has maximum value at its mean and decreases gradually on either side.

- Sometimes, it is not possible to draw the normal curve, hence it is necessary to standardize the normal curve which is known as 'standard normal curve'.

Thus 'x' is converted into 'z' and is known as 'standard normal variate'.

$$\therefore z = \frac{x-\mu}{\sigma}$$

Thus, in order to simplify the calculations standard normal variate 'z' is derived. It is possible to convert normal distribution to the standardized form because it has same shape whatever may be the values of parameters (μ and σ)

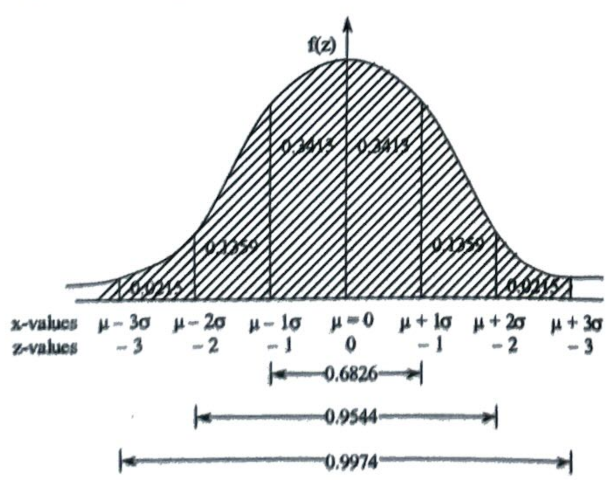


Figure: Standard Normal Distribution

5.3.1 Central Limit Theorem

Q26. Explain in detail about central limit theorem.

Answer :

Central Limit Theorem

Central Limit Theorem states that the distribution of the sum of I.I.D (Independently and Identically distributed) random variables will be normal asymptotically under general conditions with mean $\mu = \sum_{i=1}^{n} \mu_i$ and standard deviation σ where $\left(\sigma^2 = \sum_{i=1}^{n} \sigma_i^2\right)$.

Therefore the distribution $S_n = X_1 + X_2 + X_3 + \ldots X_n$ i.e., $S_n = X_i$ where S_n is the random variable whose mean $\mu_i = E(X_i)$ and variance $\sigma_i^2 = V(X_i)$.

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,

$$X_i = \begin{cases} 1 & \text{if probability is } p \\ 0 & \text{if probability is } q \end{cases} \quad \text{where } i = 1, 2, 3, \ldots n$$

Then the distribution $Sn = X_1 + X_2 + X_3 \ldots Xn$ 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 $E(X_i) = \mu_1$ and $V(X_i) = \sigma^2$ then the sum $S_n = X_1 + X_2 + X_3 + \ldots + X_n$ follows normal distribution where $\mu = n\mu_1$ (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 are not identical. In this case third absolute moment (ρ^3) is considered whose distribution can be given as,

$$\rho^3 = \sum_{i=1}^{n} \rho_i^3$$

Then under general conditions, if $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$ and $\underset{n\to\infty}{Lt} = \frac{p}{\sigma} = 0$, then the sum $X = X_1 + X_2 + X_3$ $\ldots\ldots X_n$ follows normal distribution at $N(\mu_1, \sigma^2)$ with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

PROBLEM ON CENTRAL LIMIT THEOREM

Q27. Let $X_1, X_2, \ldots$ be a I.I.D. Poisson variates with parameter λ. Use CLT to estimate $P(120 \leq S_n \leq 160)$, where $S_n = X_1 + X_2 + X_3 + \ldots + X_n$, $\lambda = 2$ and $n = 75$.

Solution :

Given that,

$X_1, X_2, X_3, \ldots, X_n$ I.I.D Poisson variates with parameter λ.

$\lambda = 2$

$n = 75$

Since given X_i's are IDD of poisson variate we have,

$E(X_i) = \lambda$ and

$\text{var}(X_i) = \lambda$

Where, $i = 1, 2, 3, \ldots\ldots n$

Therefore,

$$E(S_n) = \sum_{i=1}^{n} E(X_i)$$

$$= n.E(X_i)$$

$$= n\lambda \quad \ldots (1) \quad (\because E(X_i) = X)$$

$$\text{Variance } (S_n) = \text{Var}(X_1 + X_2 + X_3 \ldots\ldots X_n)$$

$$= \sum_{i=1}^{n} Var\ X_i$$

$$= n\ \text{Var}(X_i)$$

$$= n\lambda \quad \ldots (2) \quad (\because \text{Var}(X_i) = V)$$

Since n is large (i.e., $n = 75$) then by Lindeberg-Levy CLT we have,

$$S_n \sim N(n\lambda, n\lambda)$$

$$N(n\lambda, n\lambda)$$

$$= N(75 \times 2, 75 \times 2) \quad (\because n = 75, \lambda = 2)$$

$$= N(150, 150)$$

$$S_n = N(\mu = 150, \sigma^2 = 150) \quad (\mu = \text{Mean and } \sigma^2 = \text{Variance})$$

$\therefore$ $P(120 < S_n < 160)$ is,

$$P(120 < S_n < 160) = P\left[\frac{120-\mu}{\sqrt{\sigma}} \leq Z \leq \frac{160-\mu}{\sqrt{\sigma}}\right]$$

$$= P\left[\frac{120-150}{\sqrt{150}} \leq Z \leq \frac{160-150}{\sqrt{150}}\right] = P\left[\frac{-30}{12.24} \leq Z \leq \frac{10}{12.24}\right]$$

$$= P[-2.451 \leq Z \leq 0.816]$$

$$= P[-2.451 \leq Z \leq 0] + P[0 \leq Z \leq 0.816]$$

From the normal distribution table the value of $P[-2.451 \leq Z \leq 0] = 0.4929$ and $P[0 \leq Z \leq 0.816] = 0.2939$.

$$= 0.4929 + 0.2939$$

$$= 0.7868$$

$\therefore$ $P[120 < S_n < 160] = 0.7868$.

5.3.2 Characteristics of Normal Distribution

Q28. What is a Normal distribution? Explain the properties of Normal distribution.

Sept./Oct.-21, Q17 (OU)

OR

What is normal distribution? Write its any five features.

May/June-18, Q6(a) (KU)

Answer :

Normal Distribution

In normal distribution, the probability of occurrence of values of random variables are calculated within a range or interval whereas the probability of a particular value cannot be calculated and is always assumed to be zero. The probability distribution used mostly for dealing with quantities whose magnitude is continuously variable is normal distribution. Hence, it is called as 'Normal probability distribution'. It plays a prominent role in statistical theory and practice.

Characteristics/Properties of Normal Distribution

Following are the characteristics/features/properties of normal distribution,

1. The normal curve is 'bell-shaped' and symmetrical about the mean (skewness = 0). If the curve is folded along its central vertical axis the curves either side of the axis would coincide.
2. The height of the normal curve is maximum at its mean. Hence, the mean and mode coincide. Thus, in normal distribution, mean, mode and median are equal.
3. The height of the curve is maximum at its mean but reduces as it goes towards either of the direction but never touches the base. Hence, the curve is known as ASYMPTOTIC. The range in unlimited or infinite in both the directions.
4. The normal curve has only one mode and it known as 'unimodal' as there is only one maximum point,.
5. The points of inflexion i.e., the points where the change in curvature occurs are $\bar{x} \pm \sigma$ (or) $\mu \pm \sigma$.
6. The variables used in binomial and Poisson are discrete variables whereas normal distribution has continuous random variable.
7. The first and third quartiles are at same distance from the median.
8. The area under the normal curve is distributed as follows,
   (i) Mean $\pm$ 1 σ covers 68.27% area
   (ii) Mean $\pm$ 2 σ covers 95.45% area
   (iii) Mean $\pm$ 3 σ covers 99.73% area
9. The mean of normal distribution may be negative, zero or positive.
10. The total area under the normal curve for normal probability distribution is 1.

5.3.3 Fitting A Normal Distribution (Areas Method Only)

Q29. Write about the area under the normal curve and normal distribution.

Answer :

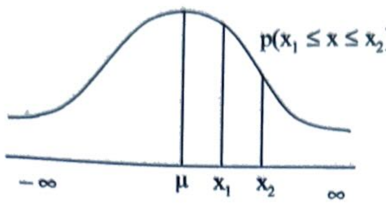
Area under any normal curve is found from the table of standard normal probability distribution showing the area between the mean and any value of the normally distributed random variable.

Since for different values of μ and σ we have different normal curves. Hence, it is not possible to draw the normal curves for various values of μ and σ. Thus, the normal curve is transformed into a standardized normal curve.

'x' is transformed into 'z' which is known as 'standard normal variate'.

$$\therefore \quad f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

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.



Figure

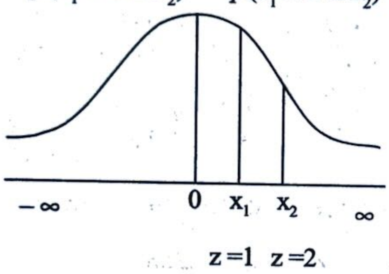
The probability can be evaluated as standardized 'z' as follows,

Substitute, $z = \frac{x-\mu}{\sigma}$ $[\mu = 0, \ \sigma = 1]$

When, $x = x_1$, $z = \frac{x_1 - \mu}{\sigma} = z_1$ (suppose)

When, $x = x_2$, $z = \frac{x_2 - \mu}{\sigma} = z_2$

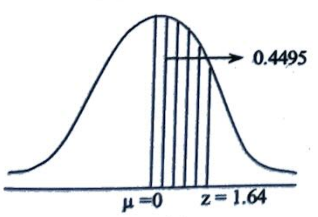
$\therefore \ p(x_1 \le x \le x_2) = p(z_1 \le z \le z_2)$



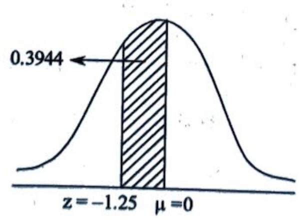
Figure

Examples

1. The area under the normal curve for $z = 1.64$
2. The area under the normal curve for $z = -1.25$



Figure



Figure

From normal distribution table, area values
If $z = 1.64$, Area = 0.4495

are taken and shade the region obtained.
$z = 1.25$, Area = 0.3944

Area Property of the Normal Distribution

The area of the standard normal curve under normal distribution is equal to one. The curve approaches horizontal axis but never touches it, hence the curve is said to be 'asymptotic'.

The area under the normal curve is split into two halves by the mean $\mu = 0$ the area on either side of the mean is 0.5.

PROBLEMS ON NORMAL DISTRIBUTION

Q30. A study of past participants indicates that the mean length of time 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 hrs

(iii) Between 550 and 650 hrs

(iv) Less than 580 hrs

(v) Between 420 and 570 hrs.

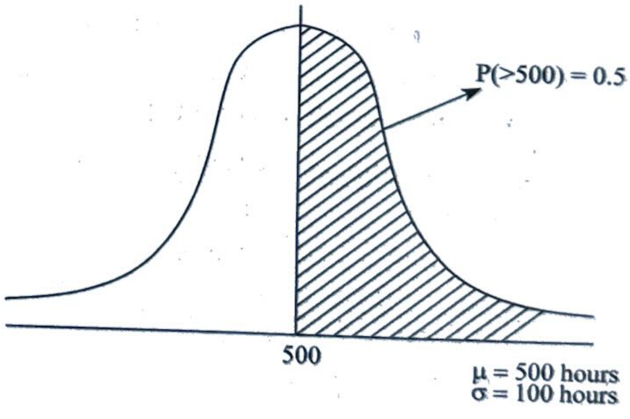
Solution : *May/June-18, Q13(b) (OU)*

Let, 'X' be length of time spent on programme (in hours) that follows normal duration.

Mean length of time spent on programme, μ = 500 hours

Standard deviation = 100 hours.

(i) More than 500 hours



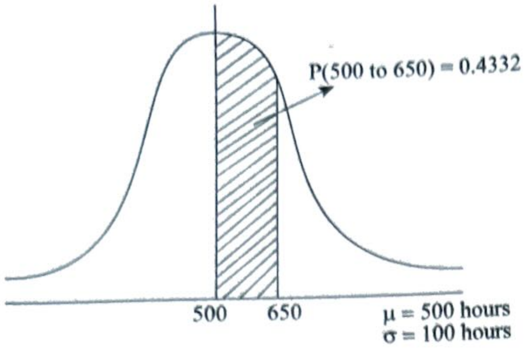
Curve's half of the area is located on either side of the mean of 500 hours. Therefore, it can be ascertain for the probability that a random variable would take on a value greater than 500 is the half shaded or 0.5.

(ii) Between 500 and 650 Hours

$P(500 \text{ to } 650)$

$Z = \frac{X-\mu}{\sigma} = \frac{650-500}{100} = \frac{150}{100}$

$= 1.5$



From the normal distribution table, the probability for $z = 1.5$ is 0.4322. Therefore, the probability that a candidate selected at random will require between 500 hours and 650 hours to complete the training program is 0.4322.

(iii) Between 550 and 650 Hours

$P(550 \text{ to } 650)$

When , $x = 550$

$$\text{Corresponding Z value, } Z = \frac{550-500}{100} = \frac{50}{100} = 0.5$$

When, $x = 650$

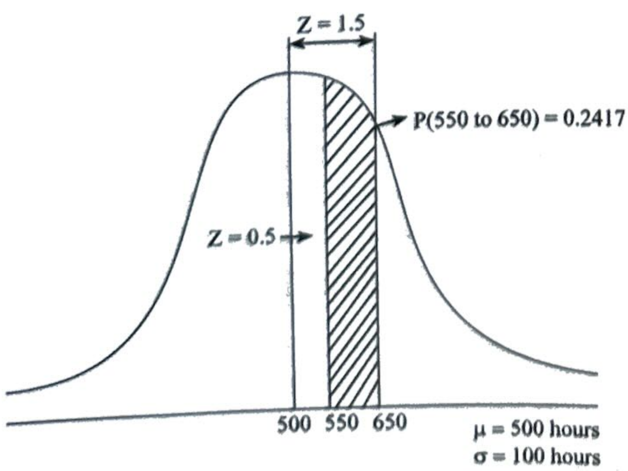
$$Z = 650 - 500 = \frac{150}{100} = 1.5$$

The area between $Z = 0.5$ and $Z = 1.5$ i.e the shaded region in the graph is shown below,

From normal distribution table,

Area at $Z = 1.5$ is 0.4332

Area at $Z = 0.5$ is 0.1915



$\therefore$ The required area $= 0.4332 - 0.1915 = 0.2417$

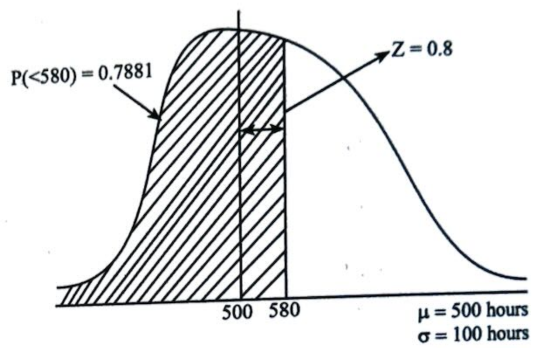
$\therefore$ The probability that a candidate selected at random will take between 550 hours and 650 hours to complete the training program is 0.2417.

(iv) Less than 580 Hours

$P(< 580)$

When,

$$x = 580, \text{ the corresponding } Z = \frac{580-500}{100} = \frac{80}{100} = 0.8$$



From normal distribution table,

Area at $Z = 0.8 = 0.2881$

The required area $= 0.2881 + 0.5$

$= 0.7881$

$\therefore$ The probability that a candidate selected at random will take lesser than 580 hours to complete the program is 0.7881.

(v) Between 420 and 570 Hours

P(420 to 570)

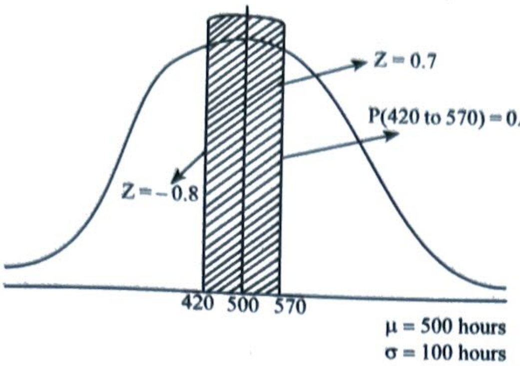
When $x = 420$

$$\text{Corresponding Z value, } Z = \frac{420-500}{100} = \frac{-80}{100} = -0.8$$

When , $x = 570$

$$Z = \frac{570-500}{100} = \frac{70}{100} = 0.7$$

The area between $Z = -0.8$ and 0.7 i.e,. the shaded region in the graph given below,



From normal distribution table,

Area at $Z = -0.8 = 0.2881$

Area at $Z = 0.7 = 0.2580$

$\therefore$ The required area $= 0.2881 + 0.2580 = 0.5461$

$\therefore$ The probability that a candidate selected at random will take between 420 hours and 570 hours to complete the training program is 0.5461.

Q31. X is normally distributed and mean and s.d of x is 12 and 4. Find out the following probabilities,

(i) $X \geq 20$

(ii) $X \leq 20$

(iii) $0 \leq X \leq 12$.

Solution :

Given that,

Mean, $\mu = 12$

Standard deviation, $\sigma = 4$

If $X \sim N(\mu, \sigma^2)$ then $Z = \dfrac{X-\mu}{\sigma}$

(i) $P(X \geq 20) = P\left(Z \geq \dfrac{X-\mu}{\sigma}\right)$

$= P\left(Z \geq \dfrac{20-12}{4}\right) = P\left(Z \geq \dfrac{8}{4}\right)$

$= P(Z \geq 2)$

$= 1 - P(Z < 2)$

$= 1 - [P(-\infty \leq Z \leq 0) + P(0 \leq Z \leq 2)]$

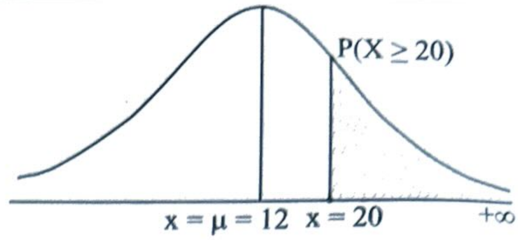
$= 1 - (0.5 + 0.4772)$

[$\because$ from table of normal distribution]

$= 1 - 0.97720$

$= 0.02280$

$\therefore P(X \geq 20) = 0.02280$



(ii) $P(X \leq 20) = P\left(Z \leq \dfrac{X-\mu}{\sigma}\right)$

$= P\left(Z \leq \dfrac{20-12}{4}\right) = P\left(Z \leq \dfrac{8}{4}\right)$

$= P(Z \leq 2)$

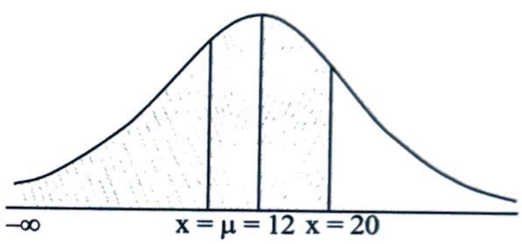
$= P(Z < 0) + P(0 < Z < 2)$

$= 0.5 + 0.4772$

[$\because$ from table of normal distribution]

$= 0.97720$

$\therefore P(X \leq 20) = 0.97720$



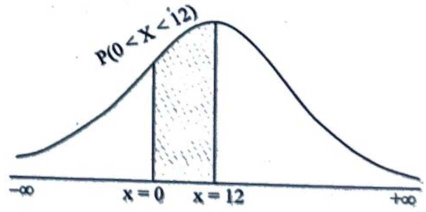
(iii) $P(0 \leq X \leq 12)$

When $X = 0$,

$$Z = \frac{X-\mu}{\sigma} = \frac{0-12}{4} = \frac{-12}{4} = -3$$

When $X = 12$,

$$Z = \frac{X-\mu}{\sigma} = \frac{12-12}{4} = \frac{0}{4} = 0$$



$P(0 \leq X \leq 12) = P(-3 < Z < 0) = P(0 < Z < 3)$

[$\because$ from table of normal distribution]

$= 0.4986$

EXERCISE AND PRACTICE QUESTIONS

SHORT QUESTIONS

THEORY

Q1. Write the Mean of Binomial Distribution. [Refer, Q1]

Q2. Bring out the differences between Binomial and Poisson distribution. [Refer, Q2]

May/June-18, Q8 (OU)

Q3. Properties of Normal Distribution. [Refer, Q5]

PROBLEMS

Q4. Comment on the following:

For a Binomial Distribution Mean = 8 and Variance = 12. [Refer Similar, Q4]

(Ans : q = 1.5).

Q5. 6 coins are tossed at a time, what is the probability of obtaining 4 or more heads?

[Refer Similar, Q7] May/June-18, Q7 (OU)

(Ans : $P(X \geq 4) = \frac{11}{32}$ *or* 0.34

Q6. What is the probability of getting 3 heads when a coin is tossed 5 times? [Refer Similar, Q7]

May/June-18, Q1(h) (KU)

(Ans : Probability of getting 3 heads when coin is tossed 5 times = $\frac{5}{16}$ or 0.3125)

Q7. A student obtained answers with mean $\mu = 2.4$ and variance $\sigma^2 = 3.2$ for a certain problem given to him using binomial distribution comment on the result. [Refer Similar, Q15]

(Ans : 1.33 (inconsistent result)).

ESSAY QUESTIONS

THEORY

Q1. What is Binomial distribution? State its importance, applications and assumptions. [Refer, Q9]

Q2. Write about the conditions and constants under which binomial distribution is used. [Refer, Q12]

Q3. Briefly describe about fitting a binomial distribution along with an illustration. [Refer, Q13]

Q4. Define poisson distribution and state its importance. [Refer, Q18]

Q5. Write about the conditions and constants under which Poisson Distribution is used. [Refer, Q20]

Q6. What is Normal Distribution? Write the importance and applications of normal distribution. [Refer, Q24]

Q7. Explain in detail about central limit theorem. [Refer, Q26]

PROBLEMS

Q8. Four coins were tossed 150 times and the following results were obtained [Refer Similar, Q15]

x	0	1	2	3	4
f	12	50	54	30	4

Fit binomial distribution under the assumption that the coins are unbiased.

(Ans : Fitted Binomial Distribution:

x	0	1	2	3	4
f	9.375	37.5	56.25	37.5	9.375

).

Q9. In an accounting department of a bank 100 accounts are selected at random and examined for errors. The following result has been obtained,

Number of Errors	0	1	2	3	4	5	6
Number of Accounts	35	40	19	2	0	2	2

Does the data verify that the errors are distributed according to the poisson probability law? [Refer Similar, Q17]

(Ans : Fitted Poisson Distribution

Number of Errors	0	1	2	3	4	5	6
Number of Accounts	35	37	19	7	2	0	0

Errors are distributed according to the Poisson's probability law).

Q10. The distribution of typing mistake committed by a typist is given below. Assuming a poisson model, find the expected frequencies. [Refer Similar, Q23]

Mistakes per page:	0	1	2	3	4	5
Number of pages:	142	156	69	27	5	1

[Ans : Fitted Poisson distribution:

Number of Mistakes per Page	0	1	2	3	4	5
Number of Pages	147	147	74	25	6	1

Q11. A random variable X is normally distributed with Mean (μ) = 12 and standard deviation (σ) = 2. Then find P(9.6 < X < 13.8) [given that $\frac{X}{\sigma} = 0.9$, A = 0.3159 and for $\frac{X}{\sigma} = 1.2$, A = 0.3849]. [Refer Similar, Q27]

[Ans : 0.7008].

Q12. A workshop produces 2000 units of an item per day. The average weight of units is 130kg with a standard deviation of 10kg. Assuming normal distribution, how many units are expected to weight less than 142kg? [Refer Similar, Q30]

[Ans : 1,770 approx].

Q13. If X is normally distributed with mean 70 and standard deviation 16. Find,

(i) $P(38 \leq X \leq 46)$

(ii) $P(82 \leq X \leq 94)$. [Refer Similar, Q31]

[Ans : $P(38 \leq x \leq 46) = 0.044$; $P(82 \leq X \leq 94) = 0.1598$].

Q14. 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. [Refer Similar, Q15]

[Ans : Mean 2.5, Standard Deviation –1.12]. May/June-18, Q13(a) (OU)

Q15. Fit a Poisson distribution to the following data: [Refer Similar, Q17]

X	0	1	2	3	4
Y	211	90	19	5	0

($e^{-m} = 0.6443$)

[Ans : $\lambda = 0.44$, Expected Frequencies –325]. May/June-19, Q13(b) (OU)

Q16. Six coins are tossed 6400 times. Find the probability to get 6 heads in 2 tosses using Poisson distribution. [Refer Similar, Q15]

[Ans : $\frac{e^{-100} . 100^2}{(100)!}$]. May/June-18, Q6(b) (KU)

INTERNAL ASSESSMENT/EXAM

I Multiple Choice

1. __________ is a discrete probability distribution developed by a Swiss mathematician. [ ]
   (a) Binomial distribution (b) Poisson distribution
   (c) Normal distribution (d) None of the above
2. Mean of the Poisson Distribution is given by __________. [ ]
   (a) np (b) p
   (c) λ (d) pq
3. __________ is also known as Bernoulli distribution. [ ]
   (a) Gamma distribution (b) Beta distribution
   (c) Normal distribution (d) Binomial distribution
4. ________ plays a major role in statistical theory and it has a widest application area than other distribution mechanism. [ ]
   (a) Rectangular distribution (b) Gamma distribution
   (c) Normal distribution (d) Beta distribution
5. The mean of binomial distribution is denoted by __________. [ ]
   (a) σ^2 (b) μ
   (c) np (d) pq
6. Poisson distribution satisfies two essential properties such as __________. [ ]
   (a) $f(x) \geq o$ (b) $\Sigma f(x) = 1$
   (c) Both (a) and (b) (d) None of the above
7. The variance of binomial distribution is __________ [ ]
   (a) $\lambda^2 = \sigma^2$ (b) $\sigma = \sqrt{\lambda}$
   (c) $\sigma^2 = npq$ (d) None of the above
8. The two parameters of normal distribution are __________. [ ]
   (a) Mean and standard deviation (b) Standard deviation and variance
   (c) Mean and variance (d) None of the above
9. Poission distribution is used generally to approximate the __________ when 'n' value is large and 'p' value is very small. [ ]
   (a) Binomial distribution (b) Gamma distribution
   (c) Normal distribution (d) None of the above
10. In a __________ mean, median and mode are all equal or they coincide with each other. [ ]
    (a) Binomial distribution (b) Beta distribution
    (c) Normal distribution (d) Gamma distribution

II Fill in the Blanks

1. Mean of Binomial distribution is __________.
2. Probability mass function of poisson distribution is __________.
3. Variance of poisson distribution is __________.
4. The probability mass function of Binomial distribution is __________.
5. The binomial distribution except that it can perform any number of trials for fixed number of success is called as __________.
6. Mean in poisson distribution is calculated by using __________ formula.
7. The sum of infinite probabilities in poisson distribution is denoted as __________.
8. Normal distribution is also known as __________.
9. The total area under the normal curve for normal probability distribution is __________.
10. Binomial distribution produces successes and failures where as poission distribution produces successes which referred as __________.

KEY

I. Multiple Choice

1. (a)
2. (c)
3. (d)
4. (c)
5. (b)
6. (c)
7. (c)
8. (c)
9. (a)
10. (c)

II. Fill in the Blanks

1. np
2. $P(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$
3. λ
4. $P(x) = {}^{n}C_{x}p^{x}q^{n-x}$
5. Negative binomial distribution
6. $\lambda = nxp$
7. P(0) + P(1) + P(2) + ,.,.,+ P(∞) = 1
8. Gaussian distribution
9. 1
10. Occurances.

III Very Short Questions and Answers

Q1. What is Binomial Distribution?

Answer :

Binomial distribution is a discrete probability distribution developed by a Swiss mathematician, 'James Bernoulli'. Binomial distribution is applicable in case of repeated trials such as,

(i) Number of applications received for a junior assistant post during a particular period of time.

(ii) Number of births taking place in a hospital.

Q2. Define Poisson Distribution.

Answer :

Poisson distribution is a discrete probability distribution for countably infinite trials. Poisson distribution is named after French mathematician, 'SIMEON DENIS POISSON' in 1837. It is used when the probability of success of any individual event is very small.

The average or mean of poisson distribution is given by λ. However, the single parameter of poisson distribution is also given as λ.

Q3. Write a short note on the conditions under which poisson distribution is used.

Answer :

Poisson distribution is a limiting case when,

(i) $n \to \infty$ i.e., number of trails is very large.

(ii) $P \to 0$ i.e., Probability of success for each trial is very small.

(iii) $np = \lambda$ is a finite constant.

Q4. Explain about central limit theorem.

Answer :

Central Limit Theorem states that the distribution of the sum of I.I.D (Independently and Identically distributed) random variables will be normal asymptotically under general conditions with mean $\mu = \sum_{i=1}^{n} \mu_i$ and standard deviation σ where $\left(\sigma^2 = \sum_{i=1}^{n} \sigma_i^2\right)$.

Q5. What do you understand by normal distribution?

Answer :

Normal distribution was first discovered by 'Abraham Demoivre' in 1733 as a limiting case of binomial distribution. It was later developed by LAPLACE and GAUSS. It is also known as 'Gaussian distribution' as the credit goes to German mathematician 'Karl Friedrich Gauss'. It is an approximation to binomial distribution whether p is equal to q or not.