GMAS601

Exam Code : MAP6A

MECHANICS AND ELEMENTS OF STATISTICS

SEMESTER - VI

MATHEMATICS

Block 2

KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

Subject Experts

Prof. N. R Das,Dept of Mathematics, G. U.

Dr. Dillip Kumar Sarma,Dept of Mathematics,Cotton College

Course Coordinator : Harekrishna Deka, KKHSOU

SLM Preparation Team

UNITS	CONTRIBUTORS
7, 8,9,10,12,13,14	Dr. Dillip Kumar Sarma Cotton College,Guwahati
11	Harekrishna Deka, KKHSOU,Guwahati

Editorial Team

Content : Dr. Dipak Kumar Sarma,Dept of Mathematics,Cotton College (Unit 11)
Harekrishna Deka , KKHSOU (Units 7, 8,9,12,13,14)

Structure & Format : Harekrishna Deka , KKHSOU

January, 2020

ISBN : 978-93-89559-99-6

© Krishna Kanta Handiqui State Open University, 2019.

This Self Learning Material (SLM) of the Krishna Kanta Handiqui State Open University is made available under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 License (international): http://creativecommons.org/licenses/by-nc-sa/4.0/

Printed and published by Registrar on behalf of the Krishna Kanta Handiqui State Open University.

Headquarters : Patgaon, Rani Gate, Guwahati-781 017; Web: www.kkhsou.in

City Office : Housefed Complex, Dispur, Guwahati-781 006

The University acknowledges with thanks the financial support provided by the **Distance Education Bureau**, New Delhi, for the preparation of this study material.

BLOCK INTRODUCTION

This is the second block of the course “Mechanics and elements of statistics”. The block consists of eight units.

In unit 7, We discuss measure of central tendency and various measures of central tendency e.g mean,median and mode.

In unit 8, We discuss measure of dispersion and various measures of dispersion e.g range,quartile deviation, mean deviation and standard deviation.

In unit 9, We discuss about moments ,kurtosis and skewness.

In unit 10,we deals with the theory of probability.We also discuss some Interesting facts about probability.

In unit 11,we discuss the concept conditional probability,Multiplication theorem of probability,Law of probability and Baye’s theorem.

In unit 12,we discuss random variable and its probability distribution.

In unit 13,we discuss binomial distribution and poisson distribution.

In unit 14,we discuss normal distribution and properties of normal distribution.

While going through a unit, you will notice some along-side boxes, which have been included to help you know some of the difficult, unseen terms. At the end of each section, you will get “CHECK YOUR PROGRESS” questions. These have been designed to self-check your progress of study. It will be better if you solve the problems put in these boxes immediately after you go through the sections of the units and then match your answers with “ANSWERS TO CHECK YOUR PPROGRESS” given at the end of each unit.

CONTENTS

		Page No.
UNIT 7:	**Measure of Central tendency** Arithmetic Mean (or Mean),Geometric Mean,Harmonic Mean,Median,Mode,Quartile	**109-129**
UNIT 8:	**Measure of Dispersion** Concept of Dispersion :Range,Quartile Deviation,Mean Deviation,Standard Deviation,Coefficient of Variation	**130-156**
UNIT 9 :	**Moments,Skewness and Kurtosis** Moments :Relation Between Rew and Central Moments Skewness :Measures of Skewness,Kurtosis:Measures of Kurtosis	**157-174**
UNIT 10 :	**Probability** Basic Terminologies,Classical Definition of Probability,Axiomatic Approach to Probability,Some Theorems on Probability	**175-197**
UNIT 11:	**Conditional probability** Conditional probability,Multiplication theorem of probability,Law of probability,Baye's theorem	**198-218**
UNIT 12 :	**Random variables and probability distribution** Random Variables : Discrete Random Variables,Continuous Random Variables, Probability Distribution of Random Variable ,Mean and Variance of Random Variables.	**219-233**
UNIT 13:	**Theoretical distribution I** Bernoulli Trial,Binomial (and Bernoulli) Distribution,Poisson DIstribution	**234-255**
UNIT 14 :	**Theoretical distribution II** Continuous Probability Distribution,Normal Distribution Properties of Normal Distribution,Standard Normal Variate	**256-270**

UNIT 7 : MEASURES OF CENTRAL TENDENCY

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Arithmetic Mean (or Mean)

7.4 Geometric Mean

7.5 Harmonic Mean

7.6 Median

7.7 Mode

7.8 Quartile

7.9 Let Us Sum Up

7.10 Further Readings

7.11 Answers To Check Your Progress

7.12 Model Questions

7.1 LEARNING OBJECTIVES

After going through this unit, you will able to learn -

- arithmetic mean
- geometric mean
- harmonic mean
- median
- mode
- quartile

7.2 INTRODUCTION

By a measure of central tendency, we mean ‘a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. There are five measures of central tendency, viz, arithmetic mean (or simply mean), geometric mean, harmonic mean, median, and mode. These measures are of common use. We will discuss these measures in this unit. Moreover we will discuss quartile which is required for our subsequent discussion.

7.3 ARITHMETIC MEAN

By the word mean, we mean arithmetic mean.

I. Mean for ungrouped data

Let there be n observations $x_1, x_2, x_3, \ldots\ldots, x_n$. The mean of these observations, denoted by $\bar{x}$, is defined by

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Note : $\sum$ is the eighteenth letter (capital) of the Greek alphabet. It is pronounced as 'sigma'.

Example 1 : The heights of 8 persons in centimetres are 155, 161, 164, 166, 167, 169, 171, 177. Find the mean height.

Solution : $\bar{x} = \frac{155+161+164+166+167+169+171+177}{8}$

$= \frac{1330}{8}$ = 166.25

$\therefore$ Mean height = 166.25 cm

II. Mean for discrete frequency distribution

Let there be n observations $x_1, x_2, \ldots., x_n$ occurring with frequencies $f_1, f_2, \ldots\ldots, f_n$ respectively. The mean of these observations, denoted by $\bar{x}$, is defined by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N}, \quad \text{Where } N = \sum_{i=1}^{n} f_i$$

Example 2 : Find the mean of the following frequency distribution.

x_i	3	5	7	9	11	13
f_i	6	8	15	25	8	4

Solution : $\bar{x} = \frac{6\times3+8\times5+15\times7+25\times9+8\times11+4\times13}{6+8+15+25+8+4}$

$$= \frac{18+40+105+225+88+52}{6+8+15+25+8+4}$$

$$= \frac{528}{66} = 8$$

III. Mean for grouped or continuous frequency distribution

Mean for grouped or continuous frequency distribution is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

where x_i is taken as the mid-value of the corresponding class.

Example 3 : Find the mean for the following frequency distribution.

Marks	0–10	10–20	20–30	30–40	40–50	50–60
No. of students	15	17	19	27	19	12

Solution : [It is a continuous frequency distribution. Upper limits of each class are excluded from the respective class. For example 10 is not included in the class 0–10; it is included in the class 10–20.]

Class interval	Frequency f_i	Mid-value x_i	$f_i x_i$
0–10	15	5	75
10–20	17	15	255
20–30	19	25	475
30–40	27	35	945
40–50	19	45	855
50–60	12	55	660

$$N = \sum f_i = 109, \quad \sum f_i x_i = 3265$$

$$\therefore \bar{x} = \frac{\sum f_i x_i}{N} = \frac{3265}{109}$$

= 29.954 (approx)

Note : In the above example f_i's and x_i's are large. So, $f_i x_i$'s are larger. In order to avoid big calculations, we use a short method to obtain the mean.

Short Method.

In this method, the mean $\bar{x}$ is given by

$$\bar{x} = a + \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a),$$

Where a is an arbitrary number, called the assumed mean.

In case of grouped or continuous frequency distribution having equal class-interval a shorter method is used. In this case, we put

$d_i = \frac{x_i - a}{h}$, where h is the width of the class-interval

The mean is given by

$$\bar{x} = a + h\frac{\sum f_i d_i}{N}$$

Example 4 : Find the mean of the frequency distribution given in Example 3 by short method.

Solution : We take a = 35. h = 10

Class interval	Frequency f_i	Mid-value x_i	$d_i = \frac{x_i - 35}{10}$	$f_i d_i$
0–10	15	5	–3	–45
10–20	17	15	–2	–34
20–30	19	25	–1	–19
30–40	27	35	0	0
40–50	19	45	1	19
50–60	12	55	2	24

$N = \sum f_i = 109$

$\sum f_i d_i = -55$

$\therefore \bar{x} = 35 + 10 \times \left(\frac{-55}{109}\right)$

$= 35 - \frac{550}{109}$

= 35 – 5.0459

= 29.9541

Propertice of Arithmetic Mean

1. The algebraic sum of the deviations of a set of values from their arithmetic mean is zero.

Proof : Let there be n observations $x_1, x_2, \ldots, x_n$ occurring with frequencies

$f_1, f_2, \ldots\ldots, f_n$ respectively. Then the mean $\bar{x}$ is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_a + f_2 + \ldots\ldots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N}, \qquad \text{where } N = \sum_{i=1}^{n} f$$

It is required to prove that $\sum_{i=1}^{n} f_i(x_i - \bar{x}) = 0$

Now, $\sum_{i=1}^{n} f_i(x_i - \bar{x})$

$= f_1(x_1 - \bar{x}) + f_2(x_2 - \bar{x}) + \ldots\ldots + f_n(x_n - \bar{x})$

$= (f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n) - (f_1 + f_2 + \ldots\ldots + f_n)\,\bar{x}$

$= \sum_{i=1}^{n} f_i x_i - \sum_{i=1}^{n} f_i\,\bar{x}$

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

= 0

2. The sum of the squares of the deviations of a set of values is minimum when taken about mean.

Proof : Let there be n observations $x_1, x_2, \ldots\ldots x_n$ occurring with frequencies $f_1, f_2, \ldots\ldots, f_n$ respectively. The mean $\bar{x}$ is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} == \frac{\sum_{i=1}^{n} f_i x_i}{N_i}, \qquad N = \sum_{i=1}^{n} f_i$$

It is required to prove that $\sum_{i=1}^{n} f_i(x_i - a)^2$ is minimum when $a = \bar{x}$, where a is an arbitrary number.

Let $z = \sum_{i=1}^{n} f_i(x_i - a)^2$

$= f_1(x_1 - a)^2 + f_2(x_2 - a)^2 + \ldots\ldots + f_n(x_n - a)^2$

$= f_1(x_1 - \bar{x} + \bar{x} - a)^2 + f_2(x_2 - \bar{x} + \bar{x} - a)^2 + \ldots\ldots + f_n(x_n - \bar{x} + \bar{x} - a)^2$

$= f_1[(x_1 - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_1 - x)(\bar{x} - a)^2]$

$+ f_2[(x_2 - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_2 - \bar{x})(\bar{x} - a)] + \ldots\ldots\ldots\ldots$

$$+ f_n[(x_n - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_n - \bar{x})(\bar{x} - a)]$$

$$= f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \ldots\ldots + f_n(x_n - \bar{x})^2 + (\bar{x} - a)^2(f_1 + f_2 + \ldots\ldots + f_n)$$
$$+ 2(\bar{x} - a)[f_1(x_1 - \bar{x}) + f_2(x_2 - \bar{x}) + \ldots\ldots + f_n(x_n - \bar{x})]$$

$$= \sum_{i=1}^{n} f_i(x_i - \bar{x})^2 + (\bar{x} - a)^2 \sum_{i=1}^{n} f_i + 2(\bar{x} - a) \sum_{i=1}^{n} f_i (x_i - \bar{x})$$

We know that $\sum_{i=1}^{n} f_i = N$ and $\sum_{i=1}^{n} f_i(x_i - \bar{x})^2 = o$ (property 1)

$$\therefore\ z = \sum_{i=1}^{n} f_i(x_i - a)^2 = \sum_{i=1}^{n} f_i(x_i - \bar{x})^2 + (\bar{x} - a)^2 N$$

If $a = \bar{x}$, $\quad (\bar{x} - a)^2 = 0$

$\therefore$ z is minimum if $a = \bar{x}$ and the minimum value is $\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$.

7.4 GEOMETRIC MEAN

Let there be n observations $x_1, x_2, \ldots\ldots, x_n$; where x_i's are non-zero and positive. The geometric mean of these observations, denoted by G, is defined by

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

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

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

$$= \frac{1}{n} \log\ [\log x_1 + \log x_2 + \ldots . + \log x_n] \qquad (1)$$

$\therefore$ G = antilog $\left[\frac{1}{n}(\log x_1 + \log x_2 + \ldots . + \log x_n)\right]$

Again, let there be n observations $x_1, x_2, \ldots\ldots, x_n$ occurring with frequencies $f_1, f_2, \ldots\ldots, f_n$ respectively. The geometric mean of these observations, denoted by G, is defined by

$$G = \left({x_1}^{f_1} {x_2}^{f_2} \ldots . {x_n}^{fn}\right)^{\frac{1}{N}}, \qquad \text{where } N = \sum_{i=1}^{n} f_i$$

$$\therefore \log G = \log \left({x_1}^{f_1} {x_2}^{f_2} \ldots . {x_n}^{fn}\right)^{\frac{1}{N}}$$

$$= \frac{1}{N}\log\left(x_1^{f_1} x_2^{f_2} \ldots . x_n^{fn}\right)$$

$$= \frac{1}{N}\left[\log x_1^{f_1} + \log x_2^{f_2} + \ldots . + \log x_n^{fn}\right]$$

$$= \frac{1}{N}\left[f_1 \log x_1 + f_2 \log x_2 + \ldots . + f_n \log x_n\right] \qquad (2)$$

$\therefore$ G = antilog $\left[\frac{1}{n}\left(f_1 \log x_1 + f_2 \log x_2 + \ldots . + f_n \log x_n\right)\right]$

From (1) and (2) we see that the logarithm of the geometric mean of n observations is equal to the arithmetic mean of the logarithms of the observations.

7.5 HARMONIC MEAN

Let there be n observations $x_1, x_2, \ldots . , x_n$, where x_i, s are non-zero.

The harmonic mean of these observations, denoted by H, is defined by

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots . + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Again, let there be in observations $x_i, x_2, \ldots . , x_n$ occurring with frequencies $f_1, f_2, \ldots . ., f_n$ respectively. The harmonic mean of these observations, denoted by H, is defined by

$$H = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots . + \frac{f_n}{x_n}} \qquad \text{where } N = \sum_{i=1}^{n} f_i$$

Example 5 : Calculate the arithmetic mean (AM), the geometric mean (GM), and the harmonic mean (HM) of 7, 8, 49, and show that AM > GM > HM.

Solution : $AM = \frac{7+8+49}{3} = \frac{64}{3} = 21.3$ (approx)

$$GM = \left(7 \times 8 \times 49\right)^{\frac{1}{3}} = \left(7 \times 2^3 \times 7^2\right)^{\frac{1}{3}} = \left(2^3 \times 7^3\right)^{\frac{1}{3}} = 2 \times 7 = 14$$

$$HM = \frac{3}{\frac{1}{7}+\frac{1}{8}+\ldots.+\frac{1}{49}} = \frac{3}{\frac{113}{392}} = 3\times\frac{392}{113} = 10.41 \text{ (approx)}$$

Clearly AM > GM > HM



CHECK YOUR PROGRESS 1

Q 1: Find the mean of the following observations 14.9, 10.8, 12.3, 23.3

Q 2: Find the mean from the following data

Marks	5	10	15	20	25	30	35	40
No. of students	5	7	9	10	8	6	3	2

Q 3: Find the mean height from the following frequency distribution.

Height (in cm)	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	42	44	58	35	26	15

Q 4: Calculate the arithmetic mean (AM), the geometric mean (GM), the harmonic mean (HM) of 7, 27, 49, and hence show that AM>GM>HM.

Q 5: Find the average speed if the first km is travelled at 10 km an hour, the next km at 8 km an hour, and the third km at 4 km an hour.

Q 6: Find the mean of the squares of the first n natural numbers.

Q 7: The numbers 3.2, 5.8, 7.9, and 4.5 have frequences x, (x + 2), (x – 3), and (x + 6) respectively. If their mean is 4.876; find the value of x.

7.6 MEDIAN

By ‘median’ of a distribution, we mean the valus of the variable which divides the distribution into two equal parts.

I. Median for ungrouped data

Let there be n observations

First we arrange the observations in ascending (or descending) order

of magnitude.

(a) Let n be odd.

In this case, the $\left(\frac{n+1}{2}\right)$th term is the median, i.e. median = $\left(\frac{n+1}{2}\right)$th term.

(b) Let n be even.

In this case, the mean of the two middle terms, i.e., the mean of the $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th terms is the median.

Note : If n is even, any value lying between the two middle values can be taken as median. It is a convention to take the mean of the two middle terms as the median.

Example 6 : Find the median of the following observations.

5, 19, 42, 11, 50, 30, 21, 0, 52, 36, 27.

Solution : We arrange the observations in ascending order of magnitude as follows. 0, 5, 11, 19, 21, 27, 30, 36, 42, 50, 52.

There are 11 observations (odd number of observations).

$\therefore$ Median = $\left(\frac{11+1}{2}\right)$th i.e., 6^{th} observation = 27

Example 7 : Find the median of the following observations

6, 10, 4, 3, 9, 11, 22, 18.

Solution : We arrange the observations in ascending order of magnitude as follows 3, 4, 6, 9, 10, 11, 18, 22.

There are 8 observations (even number of observatons).

$\therefore$ Median = Mean of the $\frac{8}{2}$th and $\left(\frac{8}{2}+1\right)$th observations

= Mean of the 4th and 5th observations

$= \frac{9+10}{2} = \frac{19}{2} = 9.5$

II. Median for discrete frequency distribution

Let there be n observations $x_1, x_2, \ldots, x_n$ with frequencies $f_1, f_2, \ldots, f_n$ respectively. In this case median is obtained by considering the cumulative frequencies.

First we find $N = \sum_{i=1}^{n} f$. Then we obtain $\frac{N}{2}$. We obtain the cumulative frequency (c.f.) just greater than $\frac{N}{2}$. The corresponding value of x is the median.

Example 8 : Find the median for the following frequency distribution.

x_i	20	9	25	50	40	80
f_i	6	4	16	7	8	2

Solution :

x_i(in anscending order)	frequency	$f_i$cumulative (c.f.)
9	4	4
29	6	10
25	16	26
40	8	34
50	7	41
80	2	43

$N = \Sigma f_i = 43$, $\frac{N}{2} = 43 = 21.5$

The c.f. just greater than 21.5 in 25. The corresponding value of x is 25. ∴ Median = 25.

III. Median for grouped or continuous frequency distribution

Median for grouped or continuous frequency distribution is given by

$$\text{Median} = L + \frac{\frac{N}{2} - c}{f} \times h$$

Where L = the lower limit of the median class,

f = the frequency of the median class,

c = the c.f. of the class preceding the median class,

h = the width of the median class.

By 'median class', we mean the class corresponding to the c.f. just

greater than $\frac{N}{2}$, where $N = \Sigma f_i$

Example 9 : Find the median for the following frequency distribution.

Class interval	20-30	30-40	40-50	50-60	60-70
Frequency	5	10	20	5	3

Solution :

Class intervel	Frequency	c.f
20-30	5	5
30-40	10	15
40-50	20	35
50-60	5	40
60-70	3	43

Here $N = \Sigma f_i = 43$, $\frac{N}{2} = \frac{43}{2} = 21.5$

The class corresponding to the c.f. just greater than 21.5 is 40-50.

$\therefore$ Median class : 40 – 50

L = 40, f = 20, c = 15, h = 20

$$\text{Median} = L + \frac{\frac{N}{2} - c}{f} \times h$$

$$= 40 + \frac{21.5 - 15}{20} \times 10$$

$$= 40 + \frac{6.5}{2}$$

= 40 + 3.25

= 43.25



CHECK YOUR PROGRESS 2

Q 1: Find the median of the following observations.

25, 30, 21, 55, 47, 10, 15, 17, 45, 35

Q 2: Find the median height from the following frequency distribution.

Heights (in cm)	150	155	158	160	164	168
No. of students	10	14	8	15	7	11

Q 3: Find the median for the following frequency distribution.

Class-interval	0-5	5-10	10-15	15-20	20-25
frequency	7	10	16	32	24

Q 4: The median of the observations 8, 11, 13, 15, x+1, x+3, 30, 35, 40, 43 arranged in ascending order is 22. Find x.

Q 5: Find the median of the following distribution.

Wages (in Rs.)	2200-2300	2300-2400	2400-2500	2500-2600	2600-2700
No. of workers	3	5	20	10	5

7.7 MODE

Mode is 'the most frequently occuring value in a distribution.'

Let us consider the following observations.

6, 5, 3, 4, 3, 7, 8, 5, 9, 5, 4.

Here the most frequently occuring value is 5.

$\therefore$ Mode of the distribution is 5.

Note : If there are more than one value each with the highest frequency, than the mode is not defined.

II. Mode for discrete frequency distribution

In case of discrete frequency distribution, mode is the value of the variable corresponding to the maximum frequency.

Let us consider the following example.

x_i	1	2	3	4	5	6	7
f_i	10	15	20	35	16	3	1

Here, maximum frequency = 35, and the corresponding value of x is 4. $\therefore$ mode = 4.

This method is not applicable in any one (or more) of the following cases.

1. The maximum frequency is repeated.
2. The maximum frequency occurs at the very beginning or at the end of

the distribution.

3. There are irregularties in the distribution.

In such cases, mode is obtained by the method of grouping. We will not discuss this method.

III. Mode for grouped or continuous frequency distribution :

Mode for grouped or continuous frequency distribution is given by

$$\text{Mode} = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h,$$

Where L = the lower limit of the modal class,

h = the width of the modal class.

f_1 = the frequency of the class preceding the modal class.

f_2 = the frequency of the class succeeding the modal class.

f_m = the frequency of the modal class.

By 'modal class', we mean the class having maximum frequency or the class in which the mode lies.

Example : Find the mode for the following distribution.

Class-interval	0-5	5-10	10-15	15-20	20-25
Frequency	7	10	16	32	24

Solution : Here, the maximum frequency is 32.

$\therefore$ the modal class is 15-20.

$\therefore$ L = 15, f_m = 32, f_1 = 16, f_2 = 24, h = 5

$$\therefore \text{Mode} = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

$$= 15 + \frac{32 - 16}{2 \times 32 - 16 - 24} \times 5$$

$$= 15 + \frac{16}{24} \times 5$$

$$= 15 + \frac{2}{3} \times 5 \quad = 15 + \frac{10}{3} = 18.3$$

Empirical relation between mean, median and mode

A distribution in which mean = median = mode is called a symmetrical distribution. If mean, median and mode are nor equal, the distribution is said to be asymmetrical or skewed. In moderately asymmetrical distribution, there exists a relationship between mean, median

and mode. Statistician karl pearson expressed the relationship as

Example: If mode=255, median=24,25 find mean.

Solotion: Mode = 3 median – 2 mean.

or, 25.5 = 3 x 24.25 – 2 mean

or, 2 mean = 72.75 – 25.5

or, 2 mean = 47.35

$\therefore$ mean $= \frac{47.25}{2}$

= 23.625



CHECK YOUR PROGRESS 3

Q 1: Find the mode of the following distribution

Heights (in cm)	0-10	10-20	20-30	30-40	40-50	50-60
No. of plants	42	44	58	35	26	15

Q 2: If mode = 16, mean = 20.2, find the median.

Q 3: If the values of mode and mean are 60, and 66 respectively, find the value of the median.

Q 4: If mode = 50.04, mean = 45, what is median?

Q 5: Find the mean, median and mode of the following observations

1, 2, 2, 3, 4, 7, 9.

7.8 QUARTILE

Quartiles are those value of the variable which divide the sequence of observations into four equal parts. Each part contains equal number of items. The first point is known as first quartile (Q_1), the second point is known as second quartle (Q_2, median), the third point is known as third quartile (Q_3).

Computation of quartiles

I. Quartiles for ungrouped data

Let there be n observations– $x_1, x_2, \ldots, x_n$.

We arrange the observations in ascending order of magnitude. Let the total number of observations be N. Then

Q_1 = Value of the $\left(\frac{N+1}{4}\right)$th observation.

Q_2 = Value of the $\left(\frac{N+1}{2}\right)$th observation, if n is odd

= mean of the values of $\frac{N}{2}$ th and $\left(\frac{N+1}{2}\right)$th observations, if n is even.

Q_3 = Value of the $3\left(\frac{N+1}{4}\right)$th observation.

II. Quartiles for grouped or continuous frequency distribution.

We compute the cumulative frequency (c.f.) of the given distribution

Let $N = \sum_{i=1}^{n} f_i$

We compute $\frac{iN}{4}$, where i = 1 for Q_1, i = 2 for Q_2 (or median), i = 3 for Q_3.

We find the cumulative frequency just greater than $\frac{iN}{4}$ and the corresponding class. The corresponding class is called the quartile class.

We use the following formula to compute Q_1, Q_2, and Q_3.

$$Q_i = L + \frac{\frac{iN}{4} - c}{f} \times h$$

i = 1, 2, 3

Where L = the lower limit of the quartile class,

f = the frequency of the quartile class,

c = the c.f. of the class preceding the quartile class,

h = the width of the quartile class.



7.9 LET US SUM UP

- Let there be n observations x_1, x_2, ..., x_n. The mean of these n observations, denoted by $\bar{x}$, is defined by

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Let there be n observations x_1, x_2, ..., x_n occurring with frequencies f_1, f_2, ..., f_n respectively. The mean of these observations is defined by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{N}$$

Mean for grouped or continuous frequency distribution is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} f_ix_i}{N}, \quad N = \sum_{i=1}^{n} f$$

where x_i taken as the mid-value of the corresponding class.

- By 'median' of a distribution, we mean the value of the variable which divides the distribution into two equal parts.

We arrange the observations in ascending (or descending) order of magnitude. If there are n observations and n is odd, the $\left(\frac{n+1}{2}\right)$th term is the median. If n is even, the mean of the $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th terms is the median.

Median for grouped or continuous frequency distribution is given by

$$\text{Median} = L + \frac{\frac{N}{2} - c}{f} \times h$$

Where L = the lower limit of the median class,

f = the frequency of the median class,

c = the c.f. of the class preceding the median class,

h = the width of the median class.

By 'median class', we mean the class corresponding to the c.f. just

greater than $\frac{N}{2}$, where $N = \Sigma f_i$

- Mode is 'the most frequently occurring value in a distribution'.

  Mode for grouped or continuous frequency distribution is given by

  $$\text{Mode} = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

Where L = the lower limit of the modal class,

h = the width of the modal class,

f_1 = the frequency of the class preceding the modal class.

f_2 = the frequency of the class succeeding the modal class,

f_m = the frequency of the modal class.

By 'modal class', we mean the class having maximum frequency or the class in which the mode lies.



7.10 FURTHER READING

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009

2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011

3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005

4) Probability, Random Process, Queseing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042

5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33



7.11 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: $\bar{x} = \frac{14.9 + 10.8 + 12.3 + 23.3}{4} = \frac{61.3}{4} = 15.325$

Ans to Q No 2:

$$\bar{x} = \frac{5\times5+7\times10+9\times15+10\times20+8\times25+6\times30+3\times35+2\times40}{5+7+9+10+8+6+3+2}$$

$$= \frac{25+70+135+200+200+180+105+80}{50}$$

$$= \frac{995}{50} = 19.9$$

Ans to Q No 3:

Class interval	frequency (f_i)	Mid-value (x_i)	$f_i x_i$
0-10	42	5	210
10-20	44	15	660
20-30	58	25	1450
30-40	35	35	1225
40-50	26	45	1170
50-60	15	55	825

$N = \Sigma f_i = 220$, $\Sigma f_i x_i = 5540$

$\therefore \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{5540}{220}$ = 25.18 (approx)

Ans to Q No 4: $AM = \frac{7+27+49}{3} = \frac{83}{3}$ = 27.67 (approx)

$$GM = (7+27+49)^{\frac{1}{3}} = (7\times3^3\times7^2)^{\frac{1}{3}} = (3^3\times7^3)^{\frac{1}{3}} = 3\times7 = 21$$

$$HM = \frac{3}{\frac{1}{7}+\frac{1}{27}+\frac{1}{49}} = \frac{3}{\frac{265}{1323}} = 3\times\frac{1323}{265}$$

$= \frac{3969}{265}$ = 14.98 (approx)

$\therefore$ AM > GM > HM

Ans to Q No 5: 1st km is travelled in $\frac{1}{10}$ hour

2nd km is travelled in $\frac{1}{8}$ hour

3rd km is travelled in $\frac{1}{4}$ hour

$\therefore$ The distance (3 km) is travelled in $\left(\frac{1}{10}+\frac{1}{8}+\frac{1}{4}\right)$ hour

$\therefore$ average speed $= \dfrac{3}{\frac{1}{10}+\frac{1}{8}+\frac{1}{4}}$ km/hour

$= \dfrac{3}{\frac{19}{40}}$ km/hour

$= 3 \times \dfrac{40}{19}$ km/hour

= 6.316 km/hour (approx)

Note : Average speed is the harmonic mean of 10, 8, 4.

Ans to Q No 6: $\bar{x} = \dfrac{1^2 + 2^2 + 3^2 + \ldots\ldots + n^2}{n}$

$= \dfrac{n(n+1)(2n+1)}{6n}$

$= \dfrac{(n+1)(2n+1)}{6}$

Ans to Q No 7: $4.876 = \dfrac{(3.2)x + (5.8)(x+2) + (7.9)(x-3) + (4.5)(x+6)}{x + x + 2 + x - 3 + x + 6}$

or, $4.876 = \dfrac{(21.4)x + 11.6 - 23.7 + 27}{4x + 5}$

or, (4.876) (4x + 5) = (21.4)x + 14.9

or, (19.504)x + 24.38 = (21.4)x + 14.9

or, (21.4)x – (19.504)x = 24.38–14.9

or, (1.896)x = 9.48

$\therefore\ x = \dfrac{9.48}{1.896} = 5$

Check Your Progress 2

Ans to Q No 1: We arrange the observations in ascending order of magnitude as follows.

10, 15, 17,21, 25, 30, 35, 45, 47, 55

There are 10 observations (even number of observations).

$\therefore$ Median = mean of the $\dfrac{10}{2}$ th and $\left(\dfrac{10}{2}+1\right)$ th observations

= mean of the 5th and 6th observations

$$= \frac{25+30}{2} = \frac{55}{2} = 27.5$$

Ans to Q No 2: Try yourself

Ans to Q No 3: Try yourself

Ans to Q No 4: The observations in ascending order are 8, 11, 13, 15, x+1, x+3, 30, 35, 40, 43.

There are 10 observations

$\therefore$ median = mean of the 5th and 6th observations

$$= \frac{x+1+x+3}{2} = \frac{2x+4}{2}$$

$= x + 2.$

Now, median = 22 (given)

$\therefore$ $x + 2 = 22$

$\therefore$ $x = 20$

Ans to Q No 5: Try yourself

Check Your Progress 3

Ans to Q No 1: Try yourself

Ans to Q No 2: Try yourself

Ans to Q No 3: Try yourself

Ans to Q No 4: Mode = 50.04, mean = 45 (given)

We know that

mode = 3 median – 2 mean

or, 50.04 = 3 median – 2 x 45

or, 3 median = 50.04 + 90

or, 3 median = 140.04

$\therefore$ Median = $\frac{140.04}{3}$ = 46.68

Ans to Q No 5: The observations are

1, 2, 2, 3, 4, 7, 9

mean $= \frac{1+2+2+3+4+7+9}{7} = \frac{28}{7} = 4$

There are 7 observations.

median $= \left(\frac{7+1}{2}\right)$ th observation

= 4th observation

= 3

mode = 2

Note : mode < median < mean.



7.12 MODEL QUESTIONS

Q 1: Let there be n observations 1, 2, 3, ..., n with frequencies 1, 2, 3, ..., n respectively. Find the arithmetic mean.

Q 2: Find the arithmetic mean (AM), the geometric mean (GM), and the harmonic mean (HM) of 2, 4, 16 and 32. Hence show that AM > GM > HM.

Q 3: The algebraic sum of the deviations of 20 observations measured from 30 is 2. Find the mean of these observations.

Q 4: The mean height of 15 students is 154 cm. It was found later on that while calculating the mean, the reading 155 cm was wrongly read as 125 cm. Find the correct mean height.

Q 5: Find the mean, the median, the mode from the following distribution.

Class-interval	0-10	10-20	20-30	30-40	40-50
frequency	3	10	17	7	6

*** ***** ***

UNIT 8 : MEASURES OF DISPERSION

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Concept of Dispersion

- 8.3.1 Range
- 8.3.2 Quartile Deviation
- 8.3.3 Mean Deviation
- 8.3.4 Standard Deviation
- 8.3.5 Coefficient of Variation

8.4 Let Us Sum Up

8.5 Further Readings

8.6 Answers To Check Your Progress

8.7 Model Questions

8.1 LEARNING OBJECTIVES

After going through this unit, you will able to learn

- concept of dispersion
- range
- quartile deviation
- mean deviation
- standard deviation
- coefficient of variation.

8.2 INTRODUCTION

Dispersion means ‘the degree of scatter of data’. Measures of dispersion provide ‘certain specific information concerning the scatter or dispersion of values in a distribution’. These measures viz. range, quartile deviation, mean deviation, standard deviation will be discussed in this unit.

8.3 CONCEPT OF DISPERSION

Let us consider the following two sets of observations.

A. 25, 26, 27, 28, 29

B. 3, 7, 13, 50, 60

In each of A, and B, there are 5 observations. The mean in each case is 27. But the observations in A are more concentrated than observations in B. Though the mean is the same, the two sets are differently constituted. The greatest variation from the mean in A is 4; whereas in B it is 35. Thus we see that the measures of central tendency do not give a clear idea of a distribution. These measues must be supplemented by other measures. Dispersion is one of such measures.

8.3.1 Range

Range is the difference between the maximum and the minimum values of the observations. Thus,

Range= M–m, where M is the maximum value and m is the minimum value.

Example 1: Find the range of the following observations.

23, 16, 14, 29, 5.6, 27, 13, 59

Solution : We arrange the data in ascending order.

13, 14, 16, 23, 27, 29, 56, 59

Here maximum value = 59

Minimum value = 13

$\therefore$ Range = 59 – 13 = 46

Note : Though range can be calculated easily, it is not a reliable measure. It is not based on the entire set of data. It depends upon the two extremes. In the above example, the two extremes are 13, and 59. Range is based upon these two.

Coefficient of range : Coefficient of range is defined as follows.

Coefficient of range $= \dfrac{M-m}{M+m}$, where M is the maximum

value, and m is the minimum value.

Example 2: Find the coefficient of range of the observations given in Example 1.

Solution : Here M = 59, m = 13

$\therefore$ Coefficient of range = $\frac{59-13}{59+13}=\frac{46}{72}=\frac{23}{36}$

8.3.2 Quartile Deviation

The difference between the third quartile and the first quartile is called interquartile range, and half the interquartile range is called quartile deviation (QD). Thus quartile deviation Q is given by

$$QD = Q = \frac{Q_3 - Q_1}{2},$$

Where Q_3 is the third quartile, and Q_1 is the first quartile.

Coefficient of quartile deviation is defined as follows.

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Example 3: Find the quartile deviation for the following data

Class-interval (Marks)	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	4	6	20	10	7	3

Solution : We know that $Q_1 = L + \frac{\frac{N}{4} - c}{f} \times h$,

Where L is the lower limit of the first quartile class,

N is the total number of frequencies,

c is the cumulative frequency of the class just above the quartile class,

f is the frequency of the quartile class,

h is the class interval.

Now,

Class interval	Number of students (f)	c.f.
0-10	4	4
10-20	6	10
20-30	20	30
40-50	7	47
50-60	3	50

$\therefore$ N = 50, $\frac{N}{4} = \frac{50}{4} = 12.5$

12.5 is contained the minimum c.f. 30.

$\therefore$ 1st quartile class : 20 – 30

L = 20, c = 10, f = 20, h = 10

$$\therefore Q_1 = 20 + \frac{12.5 - 10}{20} \times 10$$

$$= 20 + \frac{2.5}{2} = 20 + 1.25 = 21.25$$

Again, $Q_3 = L + \frac{\frac{3N}{4} - c}{f} \times h$

$$\frac{3N}{4} = \frac{3 \times 50}{4} = 37.5$$

$\therefore$ 3rd quartile class : 30 – 40

L = 30, c = 30, f = 10 h = 10

$$\therefore Q_3 = 30 + \frac{37.5 - 30}{10} \times 10$$

= 30 + 7.5 = 37.5

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

$$= \frac{37.5 - 21.25}{2}$$

$$= \frac{16.25}{2}$$

= 8.125

Note : Quartile deviation is not affected by the extreme values.



CHECK YOUR PROGRESS 1

Q 1: Write down one advantage and disadvantage of range.

Q 2: Find the range and coefficient of range of the following observations

43, 13, 26, 11, 39, 56, 7, 98, 96, 34, 65, 42 .

Q 3: Is quartile deviation affected by the extreme value ?

Q 4: What is two range of the following observations ?

6, 6, 6, 6, 6, 6, 6, 6, 6.

Q 5: Find the quartile deviation for the following data.

Class-interval	0-10	10-20	20-30	30-40	40-50
No.of students	5	8	15	16	6

8.3.3 Mean Deviation

Mean deviation is the mean of the numerical deviations from the mean, median or mode.

Mean deviation for ungrouped data

Let $x_1, x_2, x_3, \ldots\ldots x_n$ be n observations

Let $\bar{x}$ be the mean, out M be the median of the distribution.

(A) Mean deviation about the mean, denoted by MD ($\bar{x}$), is given by

$$MD(\bar{x}) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

(B) Mean deviation about the mode, denoted by MD(M), is given by

$$MD(M) = \frac{\sum_{i=1}^{n} |x_i - M|}{n}$$

Example 4: Find the mean deviation (a) about the mean, (b) about the median for the following observations

4, 5, 7, 8, 9, 13, 16, 18

Solution : (a) $\bar{x} = \frac{4+5+7+8+9+13+16+18}{8} = \frac{80}{8} = 10$

$\therefore$ the values of $(x_i - \bar{x})$ are –6, –5, –3, –2, –1, 3, 6, 8.

$\therefore$ the values of $|x_i - \bar{x}|$ are 6, 5, 3, 2, 1, 3, 6, 8.

$$\therefore MD(\bar{x}) = \frac{\sum_{i=1}^{8} |x_i - \bar{x}|}{8}$$

(b) The observations are in ascending order :

4, 5, 7, 8, 9, 13, 16, 18.

There are 8 observations

$$\text{Median } M = \frac{1}{2}\left(\frac{8}{2}\text{th observation} + \left(\frac{8}{2}+1\right)\text{th observation}\right)$$

$$= \frac{1}{2}(\text{4th observation + 5th observation})$$

$$= \frac{1}{2}\times(8+9) = \frac{17}{2} = 8.5$$

$\therefore$ the values of $(x_i - M)$ are

–4.5, –3.5, –1.5, –0.5, 0.5, 4.5, 7.5, 9.5

$\therefore$ the values of $|x_i - M|$ are

4.5, 3.5, 1.5, 0.5, 0.5, 4.5, 7.5, 9.5

$$\therefore MD(M) = \frac{\sum_{i=1}^{8} |x_i - M|}{8}$$

$$= \frac{4.5+3.5+1.5+0.5+0.5+4.5+7.5+9.5}{8}$$

$$= \frac{32}{2} = 4$$

Mean deviation for discrete frequency distribution :

Let $x_1, x_2, x_3, \ldots\ldots, x_n$ be n observations occurring with frequencies $f_1, f_2, f_3, \ldots\ldots, f_n$ respectively.

(A) Mean deviation about the mean, denoted by $MD(\bar{x})$, is given by

$$MD(\bar{x}) = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}, \quad \text{Where}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N}, \quad N = \sum_{i=1}^{n} f_i$$

(B) Mean deviation about the mode, denoted by MD (M), is given by

$$MD(M) = \frac{\sum_{i=1}^{n} f_i \, | x_i - M |}{N}$$

Example 5: Find the mean deviation (a) about the mean, (b) about the median for the following distribution

x_i	6	12	18	24	30	86
f_i	5	4	11	6	4	6

Solution :

(a)

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
6	5	30	–15	15	75
12	4	48	–9	9	36
18	11	198	–3	3	33
24	6	144	3	3	18
30	4	120	-	9	36
36	6	216	15	15	90

$$N = \sum_{i=1}^{6} f_i = 36, \quad \sum_{i=1}^{6} f_i x_i = 756$$

$$\therefore \ \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{756}{36} = 21$$

$$\therefore MD(\bar{x}) = \frac{\Sigma f_i \, | x_i - \bar{x} |}{N} = \frac{288}{36} = 8$$

(B)

x_i	f_i	cf	$x_i - M$	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
6	5	5	–12	12	60
12	4	9	–6	6	24
18	11	20	0	0	0
24	6	26	6	6	36
30	4	30	12	12	48
36	6	35	18	18	104

N = 36, an even number

$\therefore$ Median M = $\frac{1}{2}$ [$\frac{36}{2}$th observation + $\left(\frac{36}{2}+1\right)$th observation]

= $\frac{1}{2}$[18th observation +19th observation]

$= \frac{1}{2}(18+18) = 18$

$\therefore\ MD(M) = \frac{\sum_{i=1}^{6} f_i \mid x_i - M \mid}{N} = \frac{272}{36} = 7.5$

Mean deviation for continuous frequency distribution :

(A) Mean deviation about the mean, denoted by MD $(\bar{x})$, is given by

$MD(\bar{x}) = \frac{\Sigma f_i \mid x_i - \bar{x} \mid}{N}$, where $N = \Sigma f_i$

(B) Mean deviation about the median, denoted by MD (M), is given by

$$MD(M) = \frac{\Sigma f_i \mid x_i - M \mid}{N}$$

Example 6: Find the mean deviation about the mean for the following distribution.

Class-interval (Marks)	0-10	10-20	20-30	30-40	40-50	50-60
Frequency (Number of students)	4	8	14	16	4	2

Solution :

Class-interval	Frequency (f_i)	Mid point (x_i)	$f_i x_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0–10	6	5	30	–22	22	132
10–20	8	15	120	–12	12	96
20–30	14	25	350	–2	2	28
30–40	16	35	560	8	8	128
40–50	4	45	180	18	18	72
50–60	2	55	110	28	28	56

$N = \Sigma f_i = 50, \quad \Sigma f_i x_i = 1350$

$\therefore \ \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{1350}{50} = 27$

$MD(\bar{x}) = \frac{\Sigma f_i \mid x_i - \bar{x} \mid}{N}$

$= \frac{512}{50} = 10.24$

Example 7: Find the mean deviation about the median for the following distribution

Class-interval	0-10	10-20	20-30	30-40	40-50
Frequency	5	10	20	5	10

Solution :

Class-interval	frequency (f_i)	Midpoint (x_i)	c.f.	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
0–10	5	5	5	20	100
10–20	10	15	15	10	100
20–30	20	25	35	0	0
30–40	5	35	40	10	50
40–50	10	45	50	20	200

Here $N = \Sigma f_i = 50 \qquad \therefore \ \frac{N}{2} = \frac{50}{2} = 25$

The cumulative frequency just greater than 25 is 35, and the corresponding class-interval is 20–30.

$\therefore$ Median class is 20–30

Now, median $M = L + \frac{\frac{N}{2} - c}{f} \times h$

Here L = 20, $\frac{N}{2} = 25$ c = 15, f = 20, h = 10

$\therefore M = 20 + \frac{25 - 15}{20} \times 10 \quad = 20 + \frac{10}{20} \times 10 = 20 + 5 = 25$

$\therefore MD(M) = \frac{\Sigma f_i \mid x_i - M \mid}{N} = \frac{450}{50} = 9$



CHECK YOUR PROGRESS 2

Q 1: Find the mean deviation about the mean for the following observations.

1, 2, 3, 4, 5, 6, 7

Q 2: Find the mean deviation about the median for the following observations

2, 3, 4, 7, 8, 9, 10, 11, 14

Q 3: Find the mean deviation (a) about the mean, (b) about the median for the following distribution.

x_i	5	7	9	10	12	15
f_i	8	6	2	2	2	6

Q 5: The mean of 2, 4, 6, k, 8 and 15 is 7. Find the mean deviation (a) about the mean, (b) about the median.

Q 6: Find the mean deviation about the mean for the following distribution.

Class-interval	0–10	10–20	20–30	30–40	40–50
frequency	5	10	20	5	10

Q 6: Find the mean deviation about the median for the following distribution.

Class-interval	0–10	10–20	20–30	30–40	40–50
frequency	5	8	15	16	6

s

8.3.4 Standard Deviation

(A) Standard deviation for ungrouped data :

Let $x_1, x_2, x_3, \ldots, x_n$ be n observations. Let $\bar{x}$ be the mean.

The standard deviation (SD), denoted by s (Greek letter small 'sigma'), is defined by $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2}$

The square of the standard deviation is called variance. Thus,

$$\text{Variance} = s^2 = \frac{1}{n}\sum_{i=1}^{n}(xi - \bar{x})^2$$

Formula for s can be simplified as follows

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

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)$$

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

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

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

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

Thus, $SD = s = \sqrt{\frac{\Sigma\bar{x}_i^2}{n} - \bar{x}^2}$

$$\text{Variance} = s^2 = \frac{\Sigma\bar{x}_i^2}{n} - \bar{x}^2$$

Example 8: Find the SD of the first five natural numbers.

Solution : The first five natural numbers are 1, 2, 3, 4, 5.

$$\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

$$\therefore\ \bar{x}^2 = 3^2 = 9$$

$$\frac{\Sigma x_i^2}{n} = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2}{5} = \frac{55}{5} = 11$$

$$\therefore\ \frac{\Sigma x_i^2}{n} - \bar{x}^2 = 11 - 9 = 2$$

$$\therefore SD = s = \sqrt{2}$$

Example 9: The standard deviation of the observations a, 3, 6, 11 is 3.5. Find a.

Solution : We have $\bar{x} = \frac{a+3+6+11}{4} = \frac{a+20}{4}$

$$\frac{\Sigma x_i^{\,2}}{n} = \frac{a^2+3^2+6^2+11^2}{4} = \frac{a^2+9+36+121}{4} = \frac{a^2+166}{4}$$

Now $s^2 = \frac{\Sigma x_i^{\,2}}{n} - \bar{x}^2$

or, $(3.5)^2 = \frac{a^2+166}{4} - \frac{(a+20)^2}{4}$

or, $\frac{49}{4} = \frac{4(a^2+166)-(a^2+40a+400)}{16}$

or, $49 = \frac{3a^2-40a+264}{4}$

or, $3a^2 - 40a + 264 = 196$

or, $3a^2 - 40a + 68 = 0$

or, $3a^2 - 6a - 34a + 68 = 0$

or, $3a(a-2) - 34(a-2) = 0$

or, $(a-2)(3a-34) = 0$

$\therefore$ $a - 2 = 0$ or $3a - 34 = 0$

$\therefore$ $a = 2$ or $a = \frac{34}{3}$

Two properties

Property : If each of the observations $x_1, x_2, x_3, \ldots, x_n$ is increased by a non-zero constant k, the standard deviation as well as the variance remains the same.

Pf: Let the observations be $x_1, x_2, \ldots, x_n$.

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

$\therefore$ $SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Now the new observations are

$y_1 = x_1 + k, y_2 = x_2 + k, \ldots, y_n = x_n + k$

$\therefore$ new mean $\bar{y} = \frac{(x_1+k)+(x_2+k)+\ldots+(x_n+k)}{n}$

$$= \frac{(x_1+x_2+\ldots+x_n)+(k+k+\ldots+k)}{n}$$

$$= \frac{x_1 + x_2 + \ldots + x_n}{n} + \frac{nk}{n}$$

$$= \bar{x} + k$$

Now, new $SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y)^2}$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i + k - \bar{x} - k)^2}$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

= old SD.

Thus, the SD remains the same.

Property : If each of the observations $x_1, x_2, \ldots, x_n$ is multiplied by a non-zero constant k, the standard deviation is also multiplied by the same constant k.

Proof: Let the observations be $x_1, x_2, \ldots, x_n$.

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

$\therefore SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Now, the new observations are

$y_1 = kx_1, y_2 = kx_2, \ldots, y_n = kx_n$

$\therefore$ new mean $\bar{y} = \frac{kx_1 + kx_2 + \ldots + kx_n}{n}$

$$= \frac{k(x_1 + x_2 + \ldots + x_n)}{n}$$

$$= k\bar{x}$$

Now, new $SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(kx_i - k\bar{x})^2}$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}k^2(x_i - \bar{x})^2}$$

$$= \sqrt{k^2 \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= k\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

= k (old SD)

Note : If each of the observations $x_1, x_2, \ldots, x_n$ is multiplied by a non-zero constant, the variance is multiplied by k^2.

(B) Standard deviation for discrete frequency distribution

Let $x_1, x_2, \ldots, x_n$ be n observations occurring with frequencies $f_1, f_2, \ldots, f_n$ respectively.

The SD is defined by

$$SD = s = \frac{1}{n} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2, \qquad \text{where } N = \Sigma f$$

Example 10: FInd the standard deviation for the following distribution

x_i	5	15	25	35	45
f_i	5	4	6	3	2

Solution :

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
5	5	25	–16.5	272.25	136.25
15	4	60	–6.5	42.25	169
25	6	150	3.5	12.25	73.5
35	3	105	13.5	182.25	546.75
45	2	90	23.5	552.25	1104.5

Now, $N = \Sigma f_i = 20, \ \Sigma f_i x_i = 430$

$$\therefore \bar{x} = \frac{430}{20} = 21.5$$

$$\therefore \Sigma f_i (x_i - \bar{x})^2 = 3255$$

$$\therefore SD = \sqrt{\frac{3255}{20}} = \sqrt{162.75}$$

(C) Statdard deviation for continuous frequency distribution.

The standard deviation for a continuous frequency distribution is given by

$$SD = s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}$$

where $N = \sum_{i=1}^{n} f_i$, x_i is the midpoint of the i th class.

Example 11: Find the standard deviation for the following distribution.

Class-interval	0–10	10–20	20–30	30–40	40–50
Frequency	5	10	20	5	10

Solution :

Class-interval	frequency f_i	Midpoint x_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
0–10	5	5	25	–21	441	2205
10–20	10	15	150	–11	121	1210
20–30	20	25	500	–1	1	20
30–40	5	35	175	9	81	405
40–50	10	45	450	19	361	3610

Now, $N = \Sigma f_i = 50$, $\Sigma f_i x_i = 1300$

$$\therefore \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{1300}{50} = 26$$

Now, $\Sigma f_i (x_i - \bar{x})^2 = 7450$

$$\therefore SD = s = \sqrt{\frac{1}{N}\Sigma f_i (x_i - \bar{x})^2} = \sqrt{\frac{7450}{50}} = \sqrt{149}$$

Note : Standard deviation explains the average amount of variation on either side of the mean. It is considered to be the best measure of dispersion. It is widely used.

There is an empirical (based on experience) relation between quartile deviation, mean deviation, and standard deviation. It is given by

6 QD = 5 MD = 4 SD



CHECK YOUR PROGRESS 3

Q 1: The standard deviation of 5 observations is 13.25. If each of the observations is increased by 3, what is the new standard deviation?

Q 2: The standard deviation of 7 observations is 15.75. If each of the observations is multiplied by 2, what is the new standard deviation?

Q 3: The standard deviation of 5 observations is 27.45. What is the (a) quartile deviation, (b) mean deviation?

Q 4: Find the standard deviation of the following observations 13, 16, 17, 22

Q 5: Find the SD of the first n natural numbers.

Q 6: The mean and variance of the observations 1, 2, a, 6 and b are 4.4, and 8.24 respectively. Find the values of a and b.

Q 7: Find the SD for the following distribution

x_i	6	10	14	18	20
f_i	2	4	7	12	5

Q 8: Find the SD for the following distribution

Class-interval	0–10	10–20	20–30	30–40	40–50
Frequency	2	4	7	12	5

8.3.5 Coefficient of Variation

Let the mean and the standard deviation of a distribution be $\bar{x}$ and s respectively. The coefficient of variation (CV) is defined by

$$CV = \frac{s}{\bar{x}} \times 100, \quad \bar{x} \neq 0$$

This measure was introduced by statistician karl pearson.

Note : 1. CV is a unitless measure of dispersion.

2. CV is the best measure to compare the variability of two distributions.

3. A distribution with less CV is considered more consistent (less variable).
4. A distribution with more CV is considered less consistent (more variable).

Example 12: The mean and standard deviation of heights of adults (age 30 years), and children (age 10 months) are given below -

	Adults	Children
Mean	160 cm	60 cm
SD	10 cm	5 cm

Show that there is a greater variation in the heights of children than that of adults.

Solution :

$$\text{CV (adults)} = \frac{s}{\bar{x}} \times 100 = \frac{10}{160} \times 100 = 6.25$$

$$\text{CV (children)} \frac{s}{\bar{x}} \times 100 = \frac{5}{60} \times 100 = 8.33$$

Now CV (children) > CV (adults)

$\therefore$ there is greater variation in the heights of children than that of adults.



CHECK YOUR PROGRESS 4

Q 1: The mean blood pressure of 25 boys is 120 mm Hg, and SD is 10 mm Hg. The mean height and SD of the same set of boys is 160 cm and 5 cm respectively. Show that blood pressure is more variable than height.



8.4 LET US SUM UP

- Measures of dispersion provide 'certain specific information concerning the scatter or dispersion of values in a distribution'.
- Range, Quartile deviation, Mean Deviation, Standard Deviation are measures of dispersion.

- Range is the difference between the maximum and the minimum values of the observations.
- The difference between the third quartile and the first quartile is called interquartile range, and half the interquartile range is called quartile deviation. Quartile deviation is not affected by the extreme values.
- Mean deviation is the mean of the numerical deviations from the mean, median, or mode.
- The standard deviation (SD) of the observations $x_1, x_2, \ldots, x_n$; denoted by s (Greek letter small 'sigma'); is defined by

$$SD = s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- SD of observations $x_1, x_2, \ldots, x_n$ occurring with frequencies $f_1, f_2, \ldots, f_n$ respectively is defined by

$$SD = s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} fi(x_i - \bar{x})^2}, \qquad \text{where } N = \sum_{i=1}^{n} f_i$$

- Square of SD is called variance.
- Coefficient of variation (CV) is defined by $CV = \frac{s}{\bar{x}} \times 100, \quad \bar{x} \neq 0$
  Where s is the SD, and $\bar{x}$ the mean of the distribution.
- CV is a unitless measure of dispersion.
- A distribution with less CV is considered more consistent.
- A distribution with more CV is considered less consistent.



8.5 FURTHER READING

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009
2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011
3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005
4) Probability, Random Process, Queseing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042
5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33

8.6 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: Advantage : Range can be calculated easily.

Disadvantage : It is not a reliable measure. It is not based on the entire set of data. It depends upon the two extremes.

Ans to Q No 2: We arrange the data in ascending order.

7, 11, 13, 26, 34, 39, 42, 43, 56, 65, 96, 98.

Here, maximum value = 98

minimum value = 7

$\therefore$ Range = 98 – 7 = 91

Ans to Q No 3: Quartile deviation is not affected by the extreme values.

Ans to Q No 4: In this case, all the observations are identical.

Maximum value = Minimum value = 6.

$\therefore$ Range = 6 – 6 = 0

Ans to Q No 5:

Class interval	Number of students (f)	c.f.
0-10	5	5
10-20	8	13
20-30	15	28
30-40	16	44
40-50	6	50

Now, N = 50, $\frac{N}{4} = \frac{50}{4} = 12.5$

$\therefore$ 1st Quartile class : 10–20

$\therefore$ L = 10, c = 5, f = 8, h = 10

Now $Q_1 = L + \frac{\frac{N}{4} - c}{f} \times h$

$$= 10 + \frac{12.5 - 5}{8} \times 10$$

$$= 10 + 9.375 = 19.375$$

Now, $\frac{3N}{4} = \frac{3\times 50}{4} = 37.5$

$\therefore$ 3rd Quartile class : 30–40

L = 30, c = 28, f = 16, h = 10

$$Q_3 = L + \frac{\frac{3N}{4} - c}{f} \times h$$

$$= 30 + \frac{37.5 - 28}{16} \times 10$$

= 30 + 5.9375 = 35.9375

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

$$= \frac{35.9375 - 19.375}{2}$$

$$= \frac{16.5625}{2} = 8.28125$$

Check Your Progress 2

Ans to Q No 1: $\bar{x} = \frac{1+2+3+4+5+6+7}{7} = \frac{28}{7} = 4$

The values of $(x_i - \bar{x})$ are

–3, –2, –1, 0, 1, 2, 3

The values of $|x_i - \bar{x}|$ are

3, 2, 1, 0, 1, 2, 3

$$\therefore MD(\bar{x}) = \frac{\sum_{i=1}^{7} |x_i - \bar{x}|}{7} = \frac{3+2+1+0+1+2+3}{7}$$

$$= \frac{12}{7}$$

Ans to Q No 2: The observations are in ascending order

2, 3, 4, 7, 8, 9, 10, 11, 14

There are 9 observations

Median $M = \left(\frac{9+1}{2}\right)$th observation

= 5th observation = 8

The values of $(x_i - M)$ are –6, –5, –4, –1, 0, 1, 2, 3, 6

The values of $|x_i - M|$ are 6, 5, 4, 1, 0, 1, 2, 3, 6

$$\therefore MD(M) = \frac{\Sigma |x_i - M|}{9}$$

$$= \frac{6+5+4+1+0+1+2+3+6}{9}$$

$$= \frac{28}{9}$$

Ans to Q No 3: We have

(a)

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
5	8	40	–4	4	32
7	6	42	–2	2	12
9	2	18	0	0	0
10	2	20	1	1	2
12	2	24	3	3	6
15	6	90	6	6	36

$N = \Sigma f_i = 26, \ \Sigma f_i x_i = 234 \qquad \therefore \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{234}{26} = 9$

$$\therefore MD(\bar{x}) = \frac{\Sigma f_i |x_i - \bar{x}|}{N} = \frac{88}{26} = 3.38$$

(b)

x_i	f_i	cf	$x_i - M$	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
5	8	8	–2	2	16
7	6	14	0	0	0
9	2	16	2	2	4
10	2	18	3	3	6
12	2	20	5	5	10
15	6	26	8	8	48

N = 26, an even number

$$\therefore \text{Median} = \frac{1}{2}\left[\ \frac{26}{2}\text{th observation} + \left(\frac{26}{2}+1\right)\text{th observation}\ \right]$$

$$= \frac{1}{2}\left[\ 13\text{th observation} + 14\text{th observation}\ \right]$$

$$= \frac{1}{2}(7+7) = 7$$

$$MD(M) = \frac{\Sigma f_i |x_i - M|}{N} = \frac{84}{20} = 3.23$$

Ans to Q No 4: We have

$$7 = \frac{2+4+6+k+8+15}{6}$$

or, $7 = \frac{35+k}{6}$ or 42 = 35 + k, or k = 7

$\therefore$ the observations are 2, 4, 6, 7, 8, 15.

(a) We have $\bar{x} = 7$

$\therefore$ the values of $(x_i - \bar{x})$ are

–5, –3, –1, 0, 1, 8

$\therefore$ the values of $\left|x_i - \bar{x}\right|$ are

5, 3, 1, 0, 1, 8

$$\therefore MD(\bar{x}) = \frac{\Sigma |\, x_i - \bar{x}\,|}{6} = \frac{5+3+1+0+1+8}{6} = \frac{18}{6} = 3$$

(b) The observations are in ascending order :

2, 4, 6, 7, 8, 15

There are 6 observations.

$$\text{Median M} = \frac{1}{2}\left[\ \frac{6}{2}\text{ th observation} + \left(\frac{6}{2}+1\right)\text{th observation}\ \right]$$

$$= \frac{1}{2}(\text{3rd observation} + \text{4th observation})$$

$$= \frac{1}{2} \times (6+7) = \frac{13}{2} = 6.5$$

The values of $(x_i - M)$ are

–4.5, –2.5, –0.5, 0.5, 1.5, 8.5

$\therefore$ the values of $|x_i - M|$ are

4.5, 2.5, 0.5, 0.5, 1.5, 8.5

$$\therefore MD(M) = \frac{\Sigma |\, x_i - M\,|}{6} = \frac{4.5+2.5+0.5+0.5+1.5+8.5}{6}$$

$$= \frac{18}{6} = 3$$

Ans to Q No 5:

Class-interval	Frequency f_i	Mid point x_i	$f_i x_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0–10	5	5	25	–21	21	105
10–20	10	15	150	–11	11	110
20–30	20	25	500	–1	1	20
30–40	5	35	175	9	9	45
40–50	10	45	450	19	19	190

$$N = \sum f_i = 50$$
$$\sum f_i x_i = 1300$$
$$\therefore \bar{x} = \frac{\sum f_i x_i}{N} = \frac{1300}{50}$$
$$= 26$$
$$\therefore MD(\bar{x}) = \frac{\sum f_i \left| x_i - \bar{x} \right|}{N}$$
$$= \frac{470}{50}$$
$$= 9.4$$

Ans to Q No 6:

Class-interval	Frequency (f_i)	Mid point (x_i)	cf	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
0–10	6	5	5	23	115
10–20	8	15	13	13	104
20–30	15	25	28	3	45
30–40	16	35	44	7	112
40–50	6	45	50	17	102

Here $N = \Sigma f_i = 50$ $\quad \therefore \frac{N}{2} = \frac{50}{2} = 25$

The cumulative frequency just greater than 25 is 28, and the corresponding class-interval is 20–30. ∴ Median class is 20–30.

Now, median $M = L + \frac{\frac{N}{2} - c}{f} \times h$

Here L = 20, $\frac{N}{2}=25$, c = 13, f = 15, h = 10

$\therefore\ M=20+\frac{25-13}{15}\times 10\ =20+\frac{12}{15}\times 10\ =20+8=28$

$\therefore MD(M)=\frac{\Sigma f_i\,|\,x_i-M\,|}{N}=\frac{478}{50}=9.56$

Check Your Progress 3

Ans to Q No 1: We know that if each of the observations is increased by a constant, the standard deviation remains the same.

$\therefore$ the new standard deviation is also 13.25.

Ans to Q No 2: The new standard deviation = 5 x 15.75 = 78.75

Ans to Q No 3: We know that the quartile deviation, mean deviation and standard deviation are connected by 6 QD = 5 MD = 4SD

Here SD = 27.45

$\therefore$ 6 QD = 4 x 27.45

or, QD = $\frac{109.8}{6}=18.3$

Ans to Q No 4: We know that $s^2=\frac{\Sigma x_i^{\,2}}{N}-\bar{x}^2$

Now, $\bar{x}=\frac{13+16+17+22}{4}=\frac{68}{4}=17$

$\therefore\ s^2=\frac{13^2+16^2+17^2+22^2}{4}-17^2$

$=\frac{169+256+289+484}{4}-289$

$=\frac{1198}{4}-289=299.5-289=10.5$

$\therefore\ SD=s\ =\sqrt{10.5}$

Ans to Q No 5: We have $s^2=\frac{\Sigma x_i^{\,2}}{N}-\bar{x}^2$

Now $\bar{x}=\frac{1+2+3+.....+n}{n}=\frac{n(n+1)}{2n}=\frac{n+1}{2}$

$\frac{\Sigma x_i^{\,2}}{n}=\frac{1^2+2^2+3^2+.....+n^2}{n}$

$$= \frac{n(n+1)(2n+1)}{n} = \frac{(n+1)(2n+1)}{6}$$

$$\therefore\ s^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{2(n+1)(2n+1) - 3(n+1)^2}{12}$$

$$= \frac{(n+1)[2(2n+1) - 3(n+1)]}{12}$$

$$= \frac{(n+1)(4n+2-3n-3)}{12}$$

$$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$$

$$\therefore\ SD = s = \sqrt{\frac{n^2-1}{12}}$$

Ans to Q No 6: We have $\frac{1+2+a+6+b}{5} - 4.4$

or, 9 + a + b = 4.5 x 5 = 22

$\therefore$ a + b = 22 – 9 = 13 (1)

Again, $\frac{1^2+2^2+a^2+6^2+b^2}{5} = (4.4)^2 = 8.24$

or, 1 + 4 + a^2 + 36 + b^2 – 19.36 x 5 = 8.24 x 5

or, 41 + a^2 + b^2 – 96.8 = 41.2

or, a^2 + b^2 = 97 (2)

From (1), b = 13 – a

$\therefore$ from (2), a^2 + (13 – a)^2 = 97

or, a^2 + 169 – 26a + a^2 = 97

or, 2a^2 – 26a + 72 = 0

or, a^2 – 13a + 36 = 0

or, a^2 – 4a – 9a + 36 = 0

or, a(a–4) – 9(a–4) = 0

or, (a–4) (a–9) = 0

$\therefore$ a = 4, or a = 9

If a = 4, b = 13 – 4 = 9

If a = 9, b = 13 – 9 = 4

$\therefore$ the required observations are 4, 9.

Ans to Q No 7: Try yourself.

Ans to Q No 8: Try yourself.

Check Your Progress 4

Ans to Q No 1: CV (blood pressure) $= \frac{s}{\bar{x}} \times 100$

$= \frac{10}{120} \times 100 \quad = 8.3$ (approx)

CV (height) $= \frac{s}{\bar{x}} \times 100 \quad = 3.125$

$= \frac{5}{100} \times 100 \quad = 3.125$

now CV (blood pressure) > CV (height)

$\therefore$ blood pressure is more variable than height.



8.7 MODEL QUESTIONS

Q 1: In a class test 5 students obtain the following marks out of 20 in a particular subject

A	B	C	D	E
10	12	14	12	18

Calculate the standard deviation of these marks.

Q 2: Calculate the mean deviation about (a) mean, and (b) median from the following data.

Class-interval	20–30	30–40	40–50	50–60	60–70
frequency	120	201	150	75	25

Q 3: The mean and variance of 7 observations are 8 and 16 respectively. If 5 of these observations are 6, 8, 10, 12, 14; find the other two observations.

Q 4: The scores of two cricketers A and B in five matches are given below.

A :	58	59	60	54	65
B :	82	12	44	08	64

Which cricketer may be considered as more consistent?

Q 5: Coefficients of variation of two sets of observations are 58 and 69. Their standard deviations are 21.2 and 15.6 respectively. FInd their arithmetic means.

*** ***** ***

UNIT 9 : MOMENTS, SKEWNESS AND KURTOSIS

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 Moments

9.3.1 Relation Between Raw and Central Moments

9.4 Skewness

9.4.1 Measures of Skewness

9.5 Kurtosis

9.5.1 Measures of Kurtosis

9.6 Let Us Sum Up

9.7 Further Reading

9.8 Answers To Check Your Progress

9.9 Model Questions

9.1 LEARNING OBJECTIVES

After going through this unit, you will able to learn

- moments
- relation between raw moment and central moment
- skewness
- measures of skewness
- kurtosis
- measures of kurtosis.

9.2 INTRODUCTION

The notion of moments in statistics comes from the notion of moments in mechanics. In mechanics, by moment of a force we mean a force f multiplied by a perpendicular distance d, i.e., fd. The analogouy quantities in statistics are f_i, and $x_i - a$. Moment, skewness and kurtosis are important concepts of statistics.

9.3 MOMENTS

Let a variable x take the values $x_1, x_2, \ldots, x_n$ with frequencies $f_1, f_2, \ldots, f_n$ respectively.

The r th moment of x about any point a, denoted by m_r', is defined by

$$m'_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r, \quad N = \sum_{i=1}^{n} f_i \qquad (1)$$

m'_r is called the r th **raw** moment.

The r th moment of x about the mean $\bar{x}$, denoted by m_r, is defined by

$$m_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r \qquad (2)$$

m_r is called the r th **central** moment.

Note : For ungrouped data, the corresponding formulas for raw moment and central moment are given by

$$m'_r = \frac{1}{N}\sum_{i=1}^{n}(x_i - a)^r \text{ and } m_r = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^r$$

respectively.

Deductions

(A) Putting r = 0 in (1), we get

$$m'_0 = \frac{1}{N}\sum_{i=1}^{n} f_i, \quad (x_i - a)^0 = 1$$

$$= \frac{1}{N} N = 1$$

Putting r = 0 in (2), we get

$$m_0 = \frac{1}{N}\sum_{i=1}^{n} f_i = \frac{1}{N} N = 1$$

Thus, $m'_0 = m_0 = 1$, i.e. the 0 th raw as well as central moment is unity

(B) Putting r = 1 in (1), we get

$$m_1' = \frac{1}{N}\Sigma f_i(x_i - a)$$

$$= \frac{1}{N}\Sigma(f_i x_i - f_i a)$$

$$= \frac{1}{N}\left[\Sigma f_i x_i - \Sigma f_i a\right]$$

Now $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f_i x_i}{N} \Rightarrow \Sigma f_i x_i = N\bar{x}$

$\Sigma f_i a = a\Sigma f_i = aN$

$\therefore\ m_1' = \frac{1}{N}(N\bar{x} - aN) = \bar{x} - a = d,$ (say)

Thus, the 1st raw moment is $\bar{x} - a$

Putting r = 1 in (2), we get

$$m_1 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})$$

$$= \frac{1}{N}\sum_{i=1}^{n}(f_i x_i - f_i\bar{x})$$

$$= \frac{1}{N}\left(\Sigma f_i x_i - \Sigma f_i\bar{x}\right)$$

$$= \frac{1}{N}\left(N\bar{x} - N\bar{x}\right) = 0$$

Thus, the 1st central moment is zero.

(C) Putting r = 2 in (1), we get

$$m_2' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^2$$

Putting r = 2 in (2), we get

$$m_2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2 = a^2$$

Thus, the 2nd central moment is the variance.

9.3.1 Relation Between Raw and Central Moments

We have

$m_2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r$ where $N = \sum_{i=1}^{n} f_i$

We put $X_i = x_i - a,\ \ d = \bar{x} - a$

$\therefore\ X_i - d = (x_i - a) - (\bar{x} - a) = x_i - \bar{x}$

$\therefore\ m_r = \frac{1}{N}\Sigma f_i(X_i - d)^r$

$$= \frac{1}{N}\Sigma f_i\left(X_i^r - {}^rC_1 X_i^{r-1}d + {}^rC_2 X_i^{r-2}d^2 - {}^rC_3 X_i^{r-3}d^3 + \ldots\ldots + (-1)^r d^r\right)$$

$$= \frac{1}{N}\left[\Sigma f_i X_i^{\,r} - {}^{r}C_1 d\,\Sigma f_i X_i^{\,r-1} + {}^{r}C_2 d^2 \Sigma f_i X_i^{\,r-2} - {}^{r}C_3 d^3 \Sigma f_i X_i^{\,r-3} + \ldots\ldots + (-1)^r d^r \Sigma f_i\right]$$

$$= \frac{1}{N}\Sigma f_i (x_i - a)^r - {}^{r}C_1 d \frac{1}{N}\Sigma f_i (x_i - a)^{r-1} + {}^{r}C_2 d^2 \frac{1}{N}\Sigma f_i (x_i - a)^{r-2}$$

$$- {}^{r}C_3 d^3 \frac{1}{N}\Sigma f_i (x_i - a)^{r-3} + \ldots\ldots + (-1)^r d^r \frac{1}{N}\Sigma f_i$$

$$= m'_r - {}^{r}C_1 d\, m'_{r-1} + {}^{r}C_2 d^2 m'_{r-2} - {}^{r}C_3 d^3 m'_{r-3} + \ldots\ldots + (-1)^r d^r$$

$$= m'_r - {}^{r}C_1 m'_{r-1} m'_1 + {}^{r}C_2 d^2 m'_{r-2} m_1'^{2} - {}^{r}C_3 d^3 m'_{r-3} m_1'^{3} + \ldots\ldots + (-1)^r m_1'^{r} \quad (3)$$

Deductions

Putting r = 2 in (3), we get

$$m_2 = m_2' - {}^{2}C_1 m'_1 m_1' + {}^{2}C_2 m'_0 m_1'^{2}$$

$$= m_2' - 2m_1'^{2} + m_1'^{2}, \ (m'_0 = 1)$$

$$= m_2' - m_1'^{2} \qquad (4)$$

Putting r = 3 in (3), we get

$$m_3 = m_3' - {}^{3}C_1 m_2' m_1' + {}^{3}C_2 m_1' m_1'^{2} - {}^{3}C_3 m_0' m_1'^{3}$$

$$= m_3' - 3m_2' m_1' + 3m_1'^{3} - m_1'^{3}$$

$$= m_3' - 3m_2' m_1' + 2m_1'^{3} \qquad (5)$$

Putting r = 4 in (3), we get

$$m_4 = m_4' - {}^{4}C_1 m_3' m_1' + {}^{4}C_2 m_2' m_1'^{2} - {}^{4}C_3 m_1' m_1'^{3} + {}^{4}C_4 m_{4-4}' m_1'^{4}$$

$$= m_4' - 4m_3' m_1' + 6m_2' m_1'^{2} - 4m_1'^{4} + m_1'^{4}$$

$$= m_4' - 4m_3' m_1' + 6m_2' m_1'^{2} - 3m_1'^{4} \qquad (6)$$

Note : In the right hand side of (4), the coefficients are 1, –1; and their sum is 0. In the right hand side of (5), the coefficients are 1, –3, 2; and their sum is 0. In the right hand side of (6), the coefficients are 1, –4, 6, –3; and their sum is 0. In general, the sum of the coefficients in the right hand side of (3) is 0 as shown below.

$1 - {}^{r}C_1 + {}^{r}C_2 - {}^{r}C_3 + \ldots\ldots + (-1)^r = (1-1)^r = 0$

Example 1: The first moment of a variable 1 is 5. Find the mean.

Solution : We have

$$m_1' = \frac{1}{N}\Sigma f_i (x_i - 1), \ (a = 1)$$

$$= \frac{1}{N}(\Sigma f_i x_i - \Sigma f_i)$$

$$= \frac{1}{N}\Sigma f_i x_i - \frac{1}{N}\Sigma f_i, \quad \left(\frac{\Sigma f_i x_i}{N} = \bar{x},\ \Sigma f_i = N\right)$$

$$= \bar{x} - 1$$

Now, $\bar{x} - 1 = 5 \Rightarrow \bar{x} = 6$ mean = 6

Example 2: The first two moments of a variable about 2 are 1, and 16 respectively. Find the mean and the variance.

Solution : We have $m_1' = \frac{1}{N}\Sigma f_i (x_i - 2)$

$$m_2' = \frac{1}{N}\Sigma f_i (x_i - 2)^2$$

Now, $m_1' = \frac{1}{N}(\Sigma f_i x_i - 2 f_i)$

$$= \frac{\Sigma f_i x_i}{N} - \frac{2\Sigma f_i}{N}$$

$$= \bar{x} - 2$$

Now, $m_1' = 1$, (given)

$\therefore \bar{x} - 2 = 1$

$\Rightarrow \bar{x} = 3$

$\therefore$ mean = 3

Now, $m_2' = \frac{1}{N}\Sigma f_i (x_i - 2)^2$

$$= \frac{1}{N}\Sigma f_i (x_i - 3 + 1)^2$$

$$= \frac{1}{N}\Sigma f_i \left((x_i - 3)^2 + 2(x_i - 3) + 1\right)$$

$$= \frac{\Sigma f_i (x_i - 3)^2}{N} + \frac{2\Sigma f_i x_i}{N} - 5\frac{\Sigma f_i}{N}$$

Now, $m_2' = 16$, (given); and $m_2' = \frac{\Sigma f_i (x_i - 3)^2}{N}$, $(\bar{x} = 3)$

$\therefore 16 = m_2 + 2\bar{x} - 5$

or, $16 = m_2 + 2 \times 3 - 5$

$\therefore m_2 = 15$

We know that the 2nd central moment is the variance.

$\therefore$ Variance $= m_2 = 15$

Sheppard's Corrections to moments for grouped frequency distributions :

While computing moments for grouped frequency distributions, we assume that the frequencies are concentrated at the midpoints of the classes. This assumption introduces some errors. These are called grouping errors. Statistician William Flectword Sheppard (1863-1936) has developed a method for adjusting the moments, known as sheppard's corrections. Sheppard's corrections for the first four orders are given given below.

m_1' (corrected) = m_1',

m_2' (corrected) $= m_2' - \frac{h^2}{12}$,

m_3' (corrected) $= m_3' - \frac{h^2}{4} m_1'$

m_4' (corrected) $= m_4' - \frac{h^2}{2} m_2' + \frac{7}{240} h^4$

and, m_2 (corrected) $= m_2 - \frac{h^2}{12}$

m_3 (corrected) = m_3

m_4 (corrected) $= m_4 - \frac{h^2}{12} m_2 + \frac{7}{240} h^4$

where h is the width of the class-interval,

Pearson's b and g coefficients :

There are four important coefficients or moment-ratios due to Karl Pearson (1857-1936).

These are as follows

$$b_1 = \frac{m_3^2}{m_2^3}, \quad g_1 = +\sqrt{b_1}$$

$$b_2 = \frac{m_4}{m_2^2}, \quad g_2 = b_2 - 3$$

Note : These four coefficients are related to the shape of the curve. These are pure numbers.



CHECK YOUR PROGRESS 1

Q 1: The numbers of bacteria in 1ml of blood from 5 persons are 2, 3, 7, 8, 10. Find the first, second, third and fourth moments about the mean.

Q 2: The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50. Find the corresponding moments about the mean.

Q 3: The first four moments of a distribution about the value 4 of the variable are 1, 4, 10 and 45. Show that the mean is 5 and the variance is 3 and m_3 and m_4 are 0 and 26 respectively.

Q 4: The first four moments of a distribution about the value 5 of the variable are –4, 22, –117 and 560. Find the corresponding moments about zero.

Q 5: The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Find b_1, and b_2.

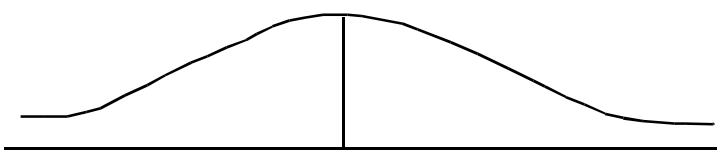
9.4 SKEWNESS

A distribution is said to be symmetrical if mean, median and mode are identical. Skewness means 'lack of symmetry'.

A symmetrical distribution when plotted on a graph gives a perfectly bell-shaped curve. If the frequency curve is not symmetrical, then the distribution is skewed.

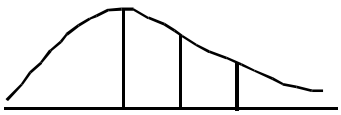
If the curve of the distribution has a longer tail towards the right, it is said to be positively skew. In this case, mean > median > mode.

If the curve of the distribution has a longer tail towards the left, it is said to be negatively skew. In this case mean < median < mode.



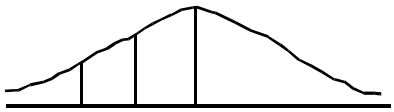
Mean = Mode = Median

Symmetrical curve



M_0 M_d M

Positively Skew



M M_d M_0

Negatively Skew

[Source : **Biostatistics** by Dr. P N Arora and Dr. P K Malhan (Himalya Publishing House, 2010)]

9.4.1 Measures of Skewness

The object of measuring skewness is to estimate (1) the extent to which, and (2) the direction in which the curve is distorted from the symmetrical curve.

The first measure of skewness = Mean – Mode

The second measure of skewness

$= (Q_3 - \text{Median}) - (\text{Median} - Q_1)$

$= Q_3 + Q_1 - 2Q_2$.

Karl Pearson's Coefficient of Skewness

We know that if a distribution is positively skew, then

mean > median > mode;

whereas if a distribution is negatively skew, then

mean < median < mode;

The difference (mean – mode) divided by the standard deviation is taken as coefficient of skewness. Thus

$$S_k = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

This is known as pearson's first coefficient of skewness.

We know that a set of data may not have a single mode. It may have two modes, called bimodal. It may have several modes, called multimodal. In this case it is said that the mode is ill defined. If the mode is ill defined, then using the empirical relation, mode = 3 median – 2 mean, for a moderately asymmetrical distribution, the

second coefficient of skewness, as suggested by Karl Pearson, is given by

$$S_k = \frac{3(\text{mean} - \text{mode})}{\text{standard deviation}}$$

Note: The above two coefficients of skewness are pure numbers, since both the numerator and denominator have the same dimension. The value of first coefficient of skewness lies between –3 and 3; and that of the second lies between –1 and 1.

Again, based upon moments, coefficient of skewness is given by

$$S_k = \frac{\sqrt{b_1}(b_2 + 3)}{2(5b_2 - 6b_1 - 9)}$$

where b_1, and b_2 have been defined earlier.

Now, if $b_1 = 0$ or $b_2 = 0$, then $S_k = 0$. But by definition $b_2 = \frac{m_4}{m_2^2}$, which cannot be negative. $\therefore b_2$ cannot be –3. Thus, $S_k = 0$ if and only if $b_1 = 0$. Thus, for a symmetrical distribution, $b_1 = 0$.

Example 3: Karl Pearson's first coefficient of skewness of a distribution's is 0.32, its standard deviation is 6.5 and the mean is 29.6. Find the mode of the distribution.

Solution : We have

$$\frac{29.6 - \text{mode}}{6.5} = 0.32$$

or, 29.6 – mode = 2.08

$\therefore$ mode = 29.6 – 2.08

i.e, mode = 27.52

9.5 KURTOSIS

The word kurtosis has been derived from the Greek word 'kurtos' meaning 'curved, arching.' If there are two distributions, they may have two same mean, same standard deviation. Also they may be equally skew. But one of them may be more peaked than the other. This measure, called

'convexity or flatness of the curve' by Karl Pearson, is known as kurtosis of the curve. It gives an idea about the peakedness of the curve.

9.5.1 Measures of Kurtosis

Kurtosis is measured by the coefficient b_2 or its derivation g_2 given by

$$b_2 = \frac{m_4}{m_2^2}, \quad g_2 = b_2 - 3$$

Note : 1. When the value of $b_2 > 3$ the curve is said to be leptokurtic.

2. When the value of $b_2 = 3$, the curve is said to be normal or mesokurtic.
3. When the value of $b_2 < 3$, the curve is said to be platykurtic.

Example 4: The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Find the kurtosis of the distribution.

Solution : We have $m_1 = 0,\ m_2 = 2.5,\ m_3 = 0.7,\ m_4 = 18.75$

Measure of kurtosis $b_2 = \frac{m_4}{m_2^2}$

$$= \frac{18.75}{(2.5)^2} = \frac{18.75}{6.25} = 3$$

Note : Since $b_2 = 3$, the curve is normal.



CHECK YOUR PROGRESS 2

Q 1: In a distribution, $Q_3 + Q_1 = 70$, $Q_3 - Q_1 = 30$, and the median is 40. Find the measure of skewness.

Q 2: The first four moments of a distribution about x = 4 are 1, 4, 10 and 45 respectively. Show the mean is 5. Find the corresponding moments about the mean. Calculate b_1 and b_2 and comment on their values.

Q 3: In a certain distribution, mean = 45, median = 48, coefficient of skewness = –0.4. Find the standard deviation.

Q 4: Karl Pearson's coefficient of skewness of a distribution is 0.4. Its standard deviation is 8, and the mean is 30. Find the mode and the median of the distribution.

Q 5: The first four central moments of a distribution are 0, 2.3, 0.9 and 15.65. Test the kurtosis of the distribution.

Q 6: Show that the following distributions have the same degree of skewness.

Distribution A : Mean 100, Median 90, SD 10

Distribution B : Mean 90, Median 80, SD 10



9.6 LET US SUM UP

- Let a variable x take the values $x_1, x_2, \ldots, x_n$ with frequencies $f_1, f_2, \ldots, f_n$ respectively.

  The r th moment of x about any point a, denoted by m_r', is defined by

  $$m_r' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^r, \quad N = \sum_{i=1}^{n} f_i$$

  m_r' is called the r th raw moment.

  The r th moment of x about the mean $\bar{x}$, denoted by m_r, is defined by

  $$m_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r$$

  m_r is called the r th central moment.

- A distribution is said to be symmetrical if mean, median and mode are identical. Skewness means 'lack of symmetry'.
- A symmetrical distribution when plotted on a graph gives a perfectly bell-shaped curve. If the frequencey curve is not symmetrical, then the distribution is skew.
- If the curve of a distribution has a longer tail towards the right, it is said to be positively skew. In this case, mean > median > mode.
- If the curve of a distribution has a longer tail towards the left, it is said to be negatively skew. In this case, mean < median < mode.
- Karl Pearson's first coefficient of skewness is given by

$$S_k = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

- Karl Pearson's second coefficient of skewness is given by

$$S_k = \frac{3(\text{mean} - \text{mode})}{\text{standard deviation}}$$

- If there are two distribution, they may have the same mean, same standard deviation. Also they may be equally skew. But one of them may be more peaked than the other. This measure, called 'convexity or flatness of the curve 'by Karl Pearson, is known as kurtosis of the curve.
- Kurtosis is measured by the coefficient b_2 or its derivation g_2 given by $b_2 = \frac{m_4}{m_2^2}$, $g_2 = b_2 - 3$



9.7 FURTHER READING

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009
2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011
3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005
4) Probability, Random Process, Queueing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042
5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33



9.8 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: Here $\bar{x} = \frac{2+3+7+8+10}{5} = 6$

We know that first moment about the mean is zero.

$\therefore\ m_1 = 0$

$$m_2 = \frac{1}{N}\Sigma(x_i - \bar{x})^2$$

$$= \frac{(2-6)^2 + (3-6)^2 + (7-6)^2 + (8-6)^2 + (10-6)^2}{5}$$

$$= \frac{16+9+1+4+16}{5} = \frac{46}{5} = 9.2$$

$$m_3 = \frac{(2-6)^3 + (3-6)^3 + (7-6)^3 + (8-6)^3 + (10-6)^3}{5}$$

$$= \frac{-64-27+1+8+64}{5} = \frac{-18}{5} = -3.6$$

$$m_4 = \frac{(2-6)^4 + (3-6)^4 + (7-6)^4 + (8-6)^4 + (10-6)^4}{5}$$

$$= \frac{256+81+1+16+256}{5} = \frac{610}{5} = 122$$

Ans to Q No 2: We have

$m_1' = 2,\ m_2' = 20,\ m_3' = 40,\ m_4' = 50$ given

Now $m_1 = 0$ (since the first moment about the mean is zero)

$$m_2 = m_2' - m_1'^2 = 20 - 4 = 16$$

$$m_3 = m_3' - 3m_2'm_1' + 2m_1'^3$$

$$= 40 - 3\times 20\times 2 + 2\times 2^3$$

$$= 40 - 120 + 16 = -64$$

$$m_4 = m_4' - 3m_3'm_1' + 6m_2'm_1'^2 - 3m_1'^4$$

$$= 50 - 4\times 40\times 2 + 6\times 20\times 2^2 - 3\times 2^4$$

$$= 50 - 320 + 480 - 48 = 162$$

Ans to Q No 3: We have $m_1' = 1,\ m_2' = 4,\ m_3' = 10,\ m_4' = 45$

Now, $m_1' = \frac{1}{N}\Sigma f_i(x_i - 4)$

or, $m_1' = \bar{x} - 4$, after simplification

$\therefore\ \bar{x} = m_1' + 4 = 1 + 4 = 5$

We know that the second central moment is the variance.

Now, $m_2 = m_2' - m_1'^2 = 4 - 1 = 3$

$\therefore$ variance = 3

$$m_3 = m_3' - 3m_2' m_1' + 2m_1'^3$$
$$= 10 - 3\times 4\times 1 + 2\times 1$$
$$= 10 - 12 + 2 = 0$$
$$m_4 = m_4' - 3m_3' m_1' + 2m_2' m_1'^2 - 3m_1'^4$$
$$= 45 - 4\times 10\times 1 + 6\times 4\times 1 - 3\times 1$$
$$= 45 - 40 + 24 - 3 = 26$$

Ans to Q No 4: We have $m_1' = -4,\ m_2' = 22,\ m_3' = -117$

$m_4' = 560,$ (given)

Now, $m_1' = \frac{1}{N}\Sigma f_i(x_i - 5)$

or, $-4 = \frac{\Sigma f_i x_i}{N} - 5\frac{\Sigma f_i}{N}$

or, $-4 = \bar{x} - 5$

$\therefore\ \bar{x} = \frac{\Sigma f_i x_i}{N} = -4 + 5 = 1$

$\therefore\ \left(m_1'\right)_0 = \frac{1}{N}\Sigma f_i(x_i - 0) = \frac{\Sigma f_i x_i}{N} = 1$

Again, $\frac{1}{N}\Sigma f_i(x_i - 5)^2 = m_2'$

or, $\frac{1}{N}\Sigma f_i(x_1^2 - 10x_i + 25)^2 = 22$

or, $\frac{1}{N}\Sigma f_i x_i^2 - 10\frac{\Sigma f_i x_i}{N} + 25\frac{\Sigma f_i}{N} = 22$

or, $\frac{1}{N}\Sigma f_i x_i^2 - 10\times 1 + 25 - 22$

or, $\frac{1}{N}\Sigma f_i x_i^2 = 22 + 10 - 25 = 7$

$\therefore\ \left(m_2'\right)_0 = \frac{1}{N}\Sigma f_i(x_i - 0)^2 = \frac{1}{N}\Sigma f_i x_i^2 = 7$

Again, $\frac{1}{N}\Sigma f_i(x_i - 5)^3 = m_3'$

or, $\frac{1}{N}\Sigma f_i(x_i^3 - 15x_i^2 + 75x_i - 125) = -117$

or, $\frac{1}{N}\Sigma f_i x_i^3 - 15\frac{\Sigma f_i x_i^2}{N} + 75\frac{\Sigma f_i x_i}{N} - 125\frac{\Sigma f_i}{N} = -117$

or, $\frac{1}{N}\Sigma f_i x_i^3 - 15\times 7 + 75\times 1 - 125 = -117$

or $\frac{1}{N}\Sigma f_i x_i^3 = 117 + 105 - 75 + 125 = 38$

$\therefore (m_3')_0 = \frac{1}{N}\Sigma f_i (x_i - 0)^3$

$= \frac{1}{N}\Sigma f_i x_i^3 = 38$

Again, $\frac{1}{N}\Sigma f_i (x_i - 5)^4 = m_4'$

or, $\frac{1}{N}\Sigma f_i (x_i^4 - 20x_i^3 + 150x_i^2 - 500x_i + 625) = 560$

or, $\frac{1}{N}\Sigma f_i x_i^4 - 20\frac{1}{N}\Sigma f_i x_i^3 + 150\frac{1}{N}\Sigma f_i x_i^2 - 500\frac{1}{N}\Sigma f_i x_i + 625\frac{1}{N}\Sigma f_i = 560$

or, $\frac{1}{N}\Sigma f_i x_i^4 - 20\times 38 + 150\times 7 - 500\times 1 + 625 = 560$

or $\frac{1}{N}\Sigma f_i x_i^4 - 760 + 1050 - 500 + 625 = 560$

$\therefore \frac{1}{N}\Sigma f_i x_i^4 = 560 + 760 - 1050 + 500 - 625 = 1820 - 1675 = 145$

$\therefore (m_4')_0 = \frac{1}{N}\Sigma f_i (x_i - 0)^4$

$= \frac{1}{N}\Sigma f_i x_i^4 = 145$

Ans to Q No 5: We have

$m_1 = 0,\ m_2 = 2.5,\ m_3 = 0.7,\ m_4 = 18.75$

Now, $b_1 = \frac{m_3^2}{m_2^3} = \frac{0.49}{15.625} = 0.03136$

$b_2 = \frac{m_4}{m_2^2} = \frac{18.75}{6.25} = 3$

Check Your Progress 2

Ans to Q No 1: We have $Q_3 + Q_1 = 70$, $Q_3 - Q_1 = 30$

$\therefore 2Q_3 = 100 \Rightarrow Q_3 = 50$

$2Q_1 = 40 \Rightarrow Q_1 = 20$

Also, median = $Q_2 = 40$.

$\therefore$ measure of skewness = $Q_3 + Q_1 - 2Q_2$

$= 50 + 20 - 2 \times 40$

$= -10$

Ans to Q No 2: We have

$m_1' = 1,\ m_2' = 4,\ m_3' = 10,\ m_4' = 5$ (given)

Now $m_1' = \frac{1}{N}\Sigma f_i(x_i - 4)$

or, $1 = \frac{\Sigma f_i x_i}{N} - 4\frac{\Sigma f_i}{N}$

or, $1 = \bar{x} - 4$

$\therefore\ \bar{x} = 1 + 4 = 6$

Now, $m_1 = 0$ (since the first moment about the mean is zero)

$m_2 = m_2' - m_1'^2 = 4 - 1 = 3$

$m_3 = m_3' - 3m_2'm_1' + 2m_1'^3$

$= 10 - 3 \times 4 \times 1 + 2 \times 1 = 10 - 12 + 2 = 0$

$m_4 = m_4' - 4m_3'm_1' + 6m_2'm_1'^2 - 3m_1'^4$

$= 45 - 4 \times 10 \times 1 + 6 \times 4 \times 1 - 3 \times 1$

$= 45 - 40 + 24 - 3$

$= 26$

Now $b_1 = \frac{m_3^2}{m_2^2} = \frac{0}{27} = 0$

$b_2 = \frac{m_4}{m_2^2} = \frac{26}{9} < 3$

Since $b_1 = 0$, the distribution is symmetrical.

Since $b_2 < 3$, the curve is platykurtic

Ans to Q No 3: mean = 45, median = 48

coefficient of skewness = – 0.4 (given)

We know that

Coefficient of skewness $= \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$

$$\therefore\ -0.4 = \frac{3\times(45-48)}{SD}$$

$$\therefore\ SD = \frac{9}{0.4} = 22.5$$

Ans to Q No 4: Coefficient of skewness = 0.4

SD = 8, mesn = 30

We know that

$$\text{Coefficient of skewness} = \frac{\text{mean} - \text{median}}{\text{SD}}$$

$$\text{or, } 0.4 = \frac{30 - \text{mode}}{8}$$

or, 3.2 = 30 – mode

$\therefore$ mode = 30 – 3.2 = 26.8

$$\text{Again, coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{SD}}$$

$$\text{or, } 0.4 = \frac{3(30 - \text{median})}{8}$$

or 3.2 = 90 – 3 (median)

$\therefore$ 3(median) = 90 – 3.2 = 86.8

$$\therefore\ \text{median} = \frac{86.8}{3} = 28.93$$

Ans to Q No 5: We have $m_1 = 0,\ m_2 = 2.3,\ m_3 = 0.9,\ m_4 = 15.65$

$$\text{Measure of kurtosis } b_2 = \frac{m_4}{m_2^2}$$

$$= \frac{15.65}{(2.3)^2}$$

$$= \frac{15.65}{5.29} = 2.9584$$

Note : Q $b_2 < 3$, the curve is platykurtic.

Ans to Q No 6: For distribution A :

$$\text{Coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{SD}}$$

$$= \frac{3\times(100-90)}{10} = 3$$

For distribution B :

$$\text{Coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{SD}}$$

$$= \frac{3\times(90-80)}{10} = 3$$

$\therefore$ distributions A and B have the same degree of skewness.



9.9 MODEL QUESTIONS

Q 1: What is meant by skewness ? What is the object of measuring skewness?

Q 2: Write short notes on

(a) Sheppard's correction to moments

(b) Skewness and kurtosis.

Q 3: Karl Pearson's coefficient of skewness of a distribution is 0.5. If mean and mode are respectively 40 and 36, find the coefficient of variation. If the median is 42, find the standard deviation.

Q 4: Karl Pearson's coefficient of skewness of a distribution is 0.32. Its standard deviation is 6.5 and mean is 29.6. Find the median and the mode of the distribution. If the mode of the above distribution is 24.8, find the standard deviation.

Q 5: The standard deviation of a symmetrical distribution is 5. What must be the value of the fourth moment about the mean in order that the distribution be (i) leptokurtic, (ii) mesokurtic and (iii) platykurtic?

*** ***** ***

UNIT 10 : PROBABILITY

UNIT STRUCTURE

10.1 Learning Objectives

10.2 Introduction

10.3 Basic Terminologies

10.4 Classical Definition of Probability

10.5 Axiomatic Approach to Probability

10.6 Some Theorems on Probability

10.6.1 Addition theorem on Probability

10.7 Let Us Sum Up

10.8 Further Reading

10.9 Answers To Check Your Progress

10.10 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, you will able to learn

- the meaning of probability
- basic terminologies of probability
- classical definition of probability
- axiomatic approach to probability
- some theorems on probability.

10.2 INTRODUCTION

'Probability is the measure of the likelihood that an event will occur.' In this unit, we will state the classical as well as the axiomatic definition of probability. We will also discuss some theorems on probability.

10.3 BASIC TERMINOLOGIES

We will discuss some basic terminologies which are required for definition of probability.

1. **Random experiment :** A random experiment is an action or an experiment which occurs under identical conditions. The result of a

random experiment is called an outcome.

Some examples of random experiment are given below.

(a) 'Tossing a coin' is a random experiment. If a fair coin is tossed, there are two possible outcomes, viz., head (H) and tail (T). Though we know that there are two possible outcomes, we cannot predict which one will occur.

(b) 'Throwing a die' is also a random experiment. A die is a solid cube, six faces of which are marked with 1, 2, 3, 4, 5, 6 dots. If a die is thrown, there are 6 possible outcomes, viz., 1, 2, 3, 4, 5, 6.

(c) 'Drawing a card from a well-shuffled pack of cards' is another example of random experiment. A pack of cards contains 52 cards in four suits, viz., Spades, Clubs, Hearts and Diamonds. Spades, and clubs are black-faced cards. On the other hand Hearts and Diamonds are red-faced cards. Each suit consists of 13 cards, of which nine cards are numbered from 2 to 10, an ace, a king, a queen, a jack (or knave). Thus, there are 4 aces, 4 kings, 4 queens and 4 jacks. Aces, kings, queens and jacks are called face cards.

2. **Trial :** The performance of a random experiment is called a trial.

3. **Event :** A collection of possible outcomes of an experiment is called an event.

4. **Equally likely events :** Events are said to be equally likely when there is no reason to expect any one of the events in preference over the others.

In the random experiment of tossing a coin, the event of getting a **Head** and the event of getting a **Tail** are equally likely events.

5. **Mutually exclusive events :** Events are said to be mutually exclusive if the occurrence of one of them rules out the occurrence of all the others.

In the random experiment of tossing a coin, the event of getting a **Head** and the event of getting a **Tail** are mutually exclusive events.

In the random experiment of drawing a card, the event of getting a

king and the event of getting a heart are not mutually exclusive events. There is a king of hearts.

6. **Exhaustive events :** Two or more events are said to be exhaustive if there is a certain chance of occurrence of at least one of them when they are all considered together. It may be noted that exhaustive events may or may not be equally likely and mutually exclusive.
7. **Favourable events :** Outcomes which entail the occurence of an event is said to be favourable to the event. In the random experiment of throwing a die, the numbers 2, 3 and 5 are favourable to the appearance of a prime number.

10.4 CLASSICAL DEFINITION OF PROBABILITY

Let there be n equally likely, mutually exclusive and exhaustive outcomes and m of them are favourable to an event E. Then the probability of the occurrence of the event E, denoted by P(E), is defined by

$$P(E) = \frac{number\ of\ outcomes\ favomable\ to\ occurence\ of\ E}{number\ of\ all\ possible\ outcomes}$$

$$= \frac{m}{n}$$

Odds of an event : Let there be m outcomes favourable to an event E, and n outcomes unfavourable to E. Then

(i) odds in favour of $E = \frac{m}{n}$

(ii) odds against $E = \frac{n}{m}$

(iii) $P(E) = \frac{m}{m+n}$

Example 1 : The odds in favour of occurrence of an event are $\frac{3}{14}$. Find the probability of the occurrence of this event.

Solution : Number of favourable outcomes = 3

Number of unfavourable outcomes = 14

$\therefore$ Total number of outcomes = 3 + 14 = 17

$\therefore$ the required probability $= \frac{3}{17}$

Example 2 : If the probability of the occurrence of an event is $\frac{6}{19}$, find the

(i) odds in favour of its occurrence,

(ii) odds against its occurrence.

Solution : Let the favourable outcomes be m and the unfavourable outcomes be n.

$\therefore$ the probability $= \frac{m}{m+n}$

But $\frac{m}{m+n} = \frac{6}{19}$, (given)

or 19m = 6m + 6n

or 13m = 6n

$\therefore \frac{m}{n} = \frac{6}{13}$

$\therefore$ (i) odds in favour $= \frac{m}{n} = \frac{6}{13}$

(ii) odds against $= \frac{n}{m} = \frac{13}{6}$

10.5 AXIOMATIC APPROACH TO PROBABILITY

The set of all possible outcomes in a random experiment is called a **sample space.** Generally it is denoted by S. Elements of a sample space are called sample points.

Two coins are flipped together. Then the outcomes are HH, HT, TH, TT. HH means head on the first coin as well as head on the second coin.HT means head on the first coin and tail on the second coin, and so on. Thus, S = {HH, HT, TH, TT}. Here the sample points are HH, HT, TH, TT, i.e. there are 4 sample points.

A subset of a sample space is called an event.

Definition : Let S be the sample space in a random experiment. Let E be an event associated with this random experiment. The probability of the

occurrence of the event E, denoted by P(E), is defined by $P(E) = \frac{n(E)}{n(S)}$

$$= \frac{number\ of\ elements\ in E}{number\ of\ elements\ in\ S}$$

Note : $P(E) = \frac{n(E)}{n(S)} \geq 0$, **Q** $n(E) \geq 0$

Example 3: Two coins are tossed together. Find the probability of getting.

(i) one head,

(ii) two heads,

(iii) at least one head,

(iv) no head.

Solution : Let S be the sample space.

$\therefore$ S = {HH, HT, TH, TT}

(i) Let E_1 be the event of getting one head

$\therefore$ E_1 = {HT, TH}

$\therefore$ $P(E_1) = \frac{n(E_1)}{n(S)} = \frac{2}{4} = \frac{1}{2}$

(ii) Let E_2 be the event of getting 2 heads.

$\therefore$ E_2 = {HH}

$\therefore$ $P(E_2) = \frac{n(E_2)}{n(S)} = \frac{1}{4}$

(iii) Let E_3 be the event of getting at least one head.

$\therefore$ E_3 = {HT, TH, HH}

$\therefore$ $P(E_3) = \frac{n(E_3)}{n(S)} = \frac{3}{4}$

(v) Let E_4 be the event of getting no head.

$\therefore$ E_4 = {TT}

$\therefore$ $P(E_4) = \frac{n(E_4)}{n(S)} = \frac{1}{4}$

Now, we discuss some terminologies again.

Impossible event : Let S be a sample space in a random experiment.

Clearly $f \subseteq S$

$\therefore$ f is an event. It is called an **impossible event.**

Example : A die is thrown.

S = { 1, 2, 3, 4, 5, 6 }, the set of all outcomes.

Let E be the event of getting a number less than 1.

Clearly $E = f$.

$\therefore$ E is an impossible event.

Note : Probability of an impossible event is zero.

Let S be a sample space in a random experiment

Let E be an impossible event, i.e. $E = f$

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{n(f)}{n(S)} = \frac{0}{n(S)} = 0$$

Sure event : Let S be a sample space in a random experiment.

Clearly $S \subseteq S$

$\therefore$ S is an event. It is called a sure event.

Example : A die is thrown.

S = { 1, 2, 3, 4, 5, 6 }

Let E be the event of getting a number less than 7.

Clearly E = 6

$\therefore$ E is a sure event.

Note : Probability of a sure event is one

Let S be a sample space in a random experiment.

Let E be a sure event, i.e., E = S

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{n(S)}{n(S)} = 1$$

Elementary and Compound events : If a sample space contains one sample point only, then it is called an elementary (or simple) event.

An event which is not elementary is called a compound (or composite) event.

Mutually exclusive events : Two events E_1 and E_2 are said to be mutually exclusive if $E_1 \cap E_2 = f$

Note : (1) $E_1 \cap E_2$ is also written as E_1E_2.

(2) If $E_1, E_2, \ldots, E_n$ are events associated with a random experiment, then $E_1, E_2, \ldots, E_n$ are mutually exclusive if $E_i \cap E_j = f$, for $i \neq j$.

Thus, E_1, E_2, E_3 are mutualy exclusive if $E_1 \cap E_2 = f$, $E_2 \cap E_3 = f$, $E_3 \cap E_1 = f$

Mutually exhaustive events : Two events E_1 and E_2 are said to be mutually exhaustive if $E_1 \cup E_2 = S$, where S is the sample space.

Note : (1) $E_1 \cup E_2$ can be denoted by $E_1 + E_2$ if $E_1 \cap E_2 = f$

(2) If $E_1, E_2, \ldots, E_n$ are events associated with a random experiment, then $E_1, E_2, \ldots, E_n$ are mutually exhaustive if $E_1 \cup E_2 \cup \ldots \cup E_n = S$ where S is the sample space.

Complementary event : Let S be a sample space, and $E \subseteq S$, i.e., E is an event. Clearly E^c is also an event, as $E^c \subseteq S$. E^c is called the complementary event. Generally it is denoted by $\overline{E}$.

Note : If w is an outcome of a trial, and $w \in E$, we say that the event E has occurred. On the other hand, if $w \notin E$, we say that the event E has not occurred

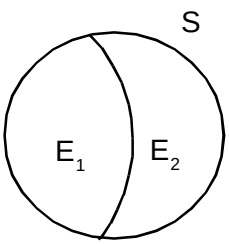
Clearly,

(i) the event E occurs if $\overline{E}$ does not occur,

(ii) the event $\overline{E}$ occurs if E does not occur.

Properties

I. Let S be a sample space. Let $E_1 \subseteq S$, $E_2 \subseteq S$ i.e., E_1, E_2 are two events. If E_1, E_2 are mutually exclusive and exhaustive events, then $P(E_1) + P(E_2) = 1$.



Proof : We have $E_1 \cap E_2 = f$

and $E_1 \cup E_2 = S$

Let n(S) = x, and $n(E_1)$ = y

$\therefore$ $n(E_2)$ = x – y, **Q** $E_1 \cap E_2 = f$, $E_2 \cup E_2 = S$

Now $P(E_1) + P(E_2) = \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)}$

$$= \frac{y}{x} + \frac{x-y}{x}$$

$$= \frac{y+x-y}{x} + \frac{x}{x} = 1$$

II. Let S be a sample space, and $E \subseteq S$, i.e., E is an event. Then $P(E) + P(\overline{E}) = 1$, i.e. $P(\overline{E}) = 1 - P(E)$

Proof : We know that E and $\overline{E}$ are mutually exclusive and exhaustive events.

$\therefore$ by the previous result, $P(E)+P(\overline{E})=1$

i.e. $P(\overline{E})=1-P(E)$

III. Let S be a sample sapce, and E_1, E_2 be two events such that $E_1 \subseteq E_2$. Then $P(E_1) \leq P(E_2)$

Proof : We have $E_1 \subseteq E_2$

$\therefore\ n(E_1) \leq n(E_2)$, ($E_1$, E_2 are finite)

or, $\dfrac{n(E_1)}{n(S)} \leq \dfrac{n(E_2)}{n(S)}$

$\therefore\ P(E_1) \leq P(E_2)$

IV. Let S be a sample space, and E be an event, i.e. $E \subseteq S$. Then $0 \leq P(E) \leq 1$

Proof : We know that

$f \subseteq E$ and $E \subseteq S$

$\therefore P(f) \leq P(E)$ and $P(E) \leq P(S)$

or, $0 \leq P(E)$ and $P(E) \leq 1$

$\therefore\ 0 \leq P(E) \leq 1$

Example 4 : Two dice are thrown. Find the probability of getting a total of 9.

Solution : Let S be the sample space

$\therefore$ S = { (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), ...

. . . . . .

. . . . . .

. . . . . .

. . . . . .

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

[Here (1, 1) means 1 in the first die, 1 in the second; (1, 2) means 1 in the first die, 2 in the second die, and so on]

$\therefore$ n (S) = 6 x 6 = 36.

Let E be the event of getting a total of 9.

$\therefore$ E = { (3, 6), 4, 5), (5, 4), (6, 3) }

$\therefore$ n (E) = 4

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{4}{36} = \frac{1}{9}$$

Example 5 : A letter of the English alphabet is chosen at random. Find the probability that the letter so chosen.

(i) is a consonant,

(ii) precedes o and is a consonant (in alphabetical order),

(iii) follows o and is a consonant.

Solution : Let S be the sample space

$\therefore$ S = {a, b, c, ..., x, y, z}

Clearly n(S) = 26

(i) Let E_1 be the event of getting a consonant. We know that 21 consonants in the English alphabet.

$$\therefore P(E_1) = \frac{n(E_1)}{n(S)} = \frac{21}{26}$$

(ii) Let E_2 be the event of getting a consonant that precedes o.

$\therefore$ E_2 = {b, c, d, f, g, h, j, k, l, m, n}

$n(E_2) = 11$

$$\therefore P(E_2) = \frac{n(E_2)}{n(S)} = \frac{11}{26}$$

(iii) Let E_3 be the event of getting a consonant that follows o.

$\therefore$ E_3 = {p, q, r, s, t, v, w, x, y, z}

$n(E_3) = 10$

$$\therefore P(E_3) = \frac{n(E_3)}{n(S)} = \frac{10}{26} = \frac{5}{13}$$

Example 6 : 4 cards are drawn at random from a pack of well-shuffled pack of 52 cards. Find the probability that 2 are kings and 2 are queens.

Solution : 4 cards can be drawn from a pack of 52 cards in $^{52}C_4$ ways.

Let S be the sample space.

$$\therefore n(S) = {}^{52}C_4$$

Let E be the event of getting 2 kings and 2 queens.

2 kings can be drawn in $^4C_2$ ways

2 queens can be drawn in $^4C_2$ ways

$\therefore$ 2 kings and 2 queens can be drawn in $^4C_2$ x $^4C_2$ ways.

$$\therefore n(E) = {}^4C_2 \times {}^4C_2$$

$$\therefore P(E) = \frac{n(E)}{n(S)}$$

$$= \frac{{}^4C_2 \times {}^4C_2}{{}^{52}C_4}$$



CHECK YOUR PROGRESS 1

Q 1: Find the probability of getting a club from a well-shuffled pack of cards.

Q 2: A bag contains 6 red, 8 black, 10 yellow, and 1 green ball. Find the probability of drawing (i) one red ball, (ii) one black ball, (iii) one white ball.

Q 3: A coin is tossed 3 times in succession. FInd the probability of getting (i) 2 heads, (ii) at least 2 heads.

Q 4: A number is chosen at random from the first 150 natural numbers. Find the probability that the number chosen is a multiple of 5.

Q 5: A single letter is selected at random the word PROBABILITY. Find the probability that it is a vowel.

Q 6: Two dice are tossed simultaneously, FInd the probability of getting a total (i) of 6, (ii) of 12, (iii) of 18.

Q 7: Two cards are drawn at random from a well-shuffled pack of 52 cards. Find the probability that one is a king and the other is a queen.

Q 8: Two cards are drawn at random from a well-shuffled pack of 52 cards. Find the probability of getting 2 hearts.

Q 9: A bag contains 4 white, 5 black, and 6 green balls. If 2 balls are drawn at random. Find the probability that none of them is black.

Q 10: A letter of the English alphabet is chosen at random. Find the probability that the letter so chosen precedes ‘h’ and is a vowel.

Q 11: What is the probability that a leap year selected at random will contain 53 Mondays?

10.6 SOME THEOREMS ON PROBABILITY

10.6.1 Addition Theorem of Probability

I. Let S be a sample space in a random experiment. Let E_1, E_2 be two events. Then

$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ and } E_2)$

i.e., $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Proof : $P(E_1 \cup E_2) = \frac{n(E_1 \cup E_2)}{n(S)}$

$$= \frac{n(E_1) + n(E_2) - (E_1 \cap E_2)}{n(S)}$$

$$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} - \frac{(E_1 \cap E_2)}{n(S)}$$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2$$

Deduction : If E_1, E_2 are mutually exclusive, then $E_1 \cap E_2 = f$ and $\therefore\ P(E_1 \cap E_2) = P(f) = 0$. Thus, for two mutually exclusive events E_1 and E_2,

$P(E_1 \cup E_2) = P(E_1) + (E_2)$

In this case we write

$P(E_1 + E_2) = P(E_1) + P(E_2)$

II. Let S be a sample space in a random experiment. Let E_1, E_2, E_3 be three events.

Then

$P(E_1 \text{ or } E_2 \text{ or } E_3)$

$= P(E_1 \cup E_2 \cup E_3)$

$= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_2 \cap E_3) - P(E_3 \cap E_1) + P(E_1 \cap E_2 \cap E_3)$

Proof : We have

$P(E_1 \cup E_2 \cup E_3)$

$= \frac{n(E_1 \cup E_2 \cup E_3)}{n(S)}$

$$= \frac{n(E_1)+n(E_2)+n(E_3)-n(E_1\cap E_2)-n(E_2\cap E_3)-n(E_3\cap E_1)+n(E_1\cap E_2\cap E_3)}{n(S)}$$

$$= \frac{n(E_1)}{n(S)}+\frac{n(E_2)}{n(S)}+\frac{n(E_3)}{n(S)}-\frac{n(E_1\cap E_2)}{n(S)}-\frac{n(E_2\cap E_3)}{n(S)}-\frac{n(E_3\cap E_1)}{n(S)}+\frac{n(E_1\cap E_2\cap E_3)}{n(S)}$$

$$= P(E_1)+P(E_2)+P(E_3)-P(E_1\cap E_2)-P(E_2\cap E_3)-P(E_3\cap E_1)+P(E_1\cap E_2\cap E_3)$$

Deduction I : If E_1, E_2, E_3 are mutually exclusive events, then $E_1 \cap E_2 = f$, $E_2 \cap E_3 = f$, $E_3 \cap E_1 = f$, $E_1 \cap E_2 \cap E_3 = f$

Thus, for three mutually exclusive events E_1, E_2 and E_3,

$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)$

i.e., $P(E_1+E_2+E_3) = P(E_1) + P(E_2) + P(E_3)$

Similarly, if E_1, E_2, E_3, ..., E_n are mutually exclusive events.

$P(E_1 + E_2 + ...+ E_n) = P(E_1) + P(E_2) + ... P(E_n)$.

II : Let E_1, E_2 be two events associated with a random experiment S. Then

(i) $P\left(\overline{E_1 \cup E_2}\right) = P\left(\overline{E_1} \cap \overline{E_2}\right)$

(ii) $P\left(\overline{E_1 \cap E_2}\right) = P\left(\overline{E_1} \cup \overline{E_2}\right)$

Proof : We have $E_1 \subseteq S$, $E_2 \subseteq S$

By De Morgan's Laws, we know that

$\left(E_1 \cup E_2\right)^c = {E_1}^c \cap {E_2}^c$

i.e., $\overline{E_1 \cup E_2} = \overline{E}_1 \cap \overline{E}_2$

$\therefore P\left(\overline{E_1 \cup E_2}\right) = P\left(\overline{E}_1 \cap \overline{E}_2\right)$

Also $\left(E_1 \cap E_2\right)^c = {E_1}^c \cup {E_2}^c$

i.e. $\left(\overline{E_1 \cap E_2}\right) = \overline{E}_1 \cap \overline{E}_2$

$\therefore P\left(\overline{E_1 \cap E_2}\right) = \left(\overline{E}_1 \cap \overline{E}_2\right)$

Example 7 : Two dice are tossed together. Find the probability of getting a doublet or a total of 8.

Solution : Let S be the sample space.

Clearly n(S) = 36 [See Example 4]

Let E_1 be the event of getting a doublet. [Doublet means 'either of two dice, each of which, when thrown, has the same number of spots on the face lying uppermost'.]

$\therefore$ E_1 = { (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6) }

Let E_2 be the event of getting a total of 8.

$\therefore$ E_2 = { (2, 6), (3, 5), (4, 4), (5, 3), (6, 2) }

$E_1 \cap E_2 = \{(4, 4)\}$

Thus $n(E_1) = 6$, $n(E_2) = 5$, $n(E_1 \cap E_2) = 1$

$\therefore$ P(E_1 or E_2)

$$= P(E_1 \cup E_2)$$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} - \frac{n(E_1 \cap E_2)}{n(S)}$$

$$= \frac{6}{36} + \frac{5}{36} - \frac{1}{36}$$

$$= \frac{6+5-1}{36}$$

$$= \frac{10}{36}$$

$$= \frac{5}{18}$$

Example 8 : Two dice are tossed. Find the probability of getting an odd number in the first die or a total of 8.

Solution : Let S be the sample space

$\therefore$ n(S) = 36

Let E_1 be the event of getting an odd number in the first die.

$\therefore$ $n(E_1) = 3 \times 6 = 18$

[There are 3 odd numbers, viz., 1, 3, 5. Any one of them will appear in the first die. Any one of 1, 2, 3, 4, 5, 6 will appear in the second die.]

Let E_2 be the event of getting a total of 8.

$\therefore$ E_2 = {(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)}

$n(E_2) = 5$

$E_1 \cap E_2 = \{(3, 5), (5, 7)\}$

$n(E_1 \cap E_2) = 2$

$\therefore P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

$$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} - \frac{n(E_1 \cap E_2)}{n(S)}$$

$$= \frac{18}{36} + \frac{5}{36} - \frac{2}{36}$$

$$= \frac{18+5-2}{36} = \frac{21}{36} = \frac{7}{12}$$

Example 9 : Let E_1, E_2, E_3 be three mutually exclusive and exhaustive events associated with a random experiment. If $P(E_1) = \frac{3}{2}P(E_2)$ and $P(E_3) = \frac{1}{2}P(E_2)$, find $P(E_1)$, $P(E_2)$, and $P(E_3)$

Solution : Let S be the sample spare

$\therefore E_1 \cup E_2 \cup E_3 = S$ (**Q** E_1, E_2, E_3 are exhaustive events)

$\therefore P(E_1 \cup E_2 \cup E_3) = P(S)$

or, $P(E_1) + P(E_2) + P(E_3) = P(S)$ (**Q** E_1, E_2, E_3 are mutually exclusive events)

or, $\frac{3}{2}P(E_2) + P(E_2) + \frac{1}{2}P(E_2) = 1$

or, $\left(\frac{3}{2} + 1 + \frac{1}{2}\right)P(E_2) = 1$

or, $3P(E_2) = 1$

$\therefore P(E_2) = \frac{1}{3}$

$\therefore P(E_1) = \frac{3}{2}.P(E_2) = \frac{3}{2}.\frac{1}{3} = \frac{1}{2}$

$P(E_3) = \frac{1}{2}.P(E_2) = \frac{1}{2}.\frac{1}{3} = \frac{1}{6}$

Example 10 : Let E_1, E_2, be two events associated with a random experiment.

If $P(E_1) = \frac{1}{2}$, $P(E_2) = \frac{1}{3}$ and $P(E_1 \cap E_2) = \frac{1}{6}$, find

(i) $P\left(\overline{E_1 \cap E_2}\right)$, (ii) $P\left(\overline{E_1 \cup E_2}\right)$

Solution : (i) $P\left(\overline{E_1 \cap E_2}\right)$

$= 1 - p\left(E_1 \cap E_2\right)$ $[P(\bar{E}) = 1 - P(E)]$

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

(ii) $P\left(\overline{E_1 \cup E_2}\right)$

$= 1 - P(E_1 \cup E_2)$

$= 1 - [P(E_1) + P(E_2) - P(E_1 \cap E_2)]$

$= 1 - \left(\frac{1}{2} + \frac{1}{3} - \frac{1}{6}\right) = 1 - \frac{2}{3} = \frac{1}{3}$



CHECK YOUR PROGRESS 2

Q 1: A number is chosen at random from the first 150 natural numbers. Find the probability that the number so chosen is a multiple of 4 or a multiple of 5.

Q 2: A card is drawn at random from a well-shuffled pack of 52 cards. Find the probability that the card so chosen is a spade or a king.

Q 3: Let E_1, E_2 be two events associated with a random experiment. If $P(E_1) = 0.3$, $P(E_2) = 0.2$, and $P(E_1 \cap E_2) = 0.1$, Find

(i) $P\left(\overline{E_1 \cap E_2}\right)$, (ii) $P\left(\overline{E_1 \cup E_2}\right)$

Q 4: Let E_1, E_2 be two mutually exclusive events associated with a random experiment. If P(not E_1) = 0.45 and P(E_1 or E_2) = 0.65, find P(E_2).

Q 5: Let E_1, E_2 be two events associated with a random experiment. If $P(E_1) = \frac{1}{2}$, $P(E_2) = \frac{1}{3}$, P(E_1 and E_2) $= \frac{1}{6}$, Find (i) P(E_1 or E_2), (ii) P (neither E_1 nor E_2)

Q 6: Two dice are thrown together. Find the probability that the sum of the numbers on the two faces is neither divisible by 3 nor by 4.



10.7 LET US SUM UP

I A random experiment is an action or an experiment which occurs under identical conditions. The result of a random experiment is called an outcome. 'Tossing of a coin', 'Throwing a die', 'Drawing a card from a

well-shuffled pack of cards' are examples of random experiment.

- The set of all possible outcomes in a random experiment is called a sample space.
- A subset of a sample space is called an event.
- The probability of the occurrence of the event E, denoted by P(E), is defined by $P(E) = \frac{n(E)}{n(S)}$

  Where n(E) denotes the number of elements in the event E, and n(S) denotes the number of elements in the sample space S.
- Let S be a sample space in a random experiment. Let E_1, E_2 be two events. Then

  $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap_2)$



10.8 FURTHER READING

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009
2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011
3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005
4) Probability, Random Process, Queueing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042
5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33



10.9 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: Let S be the sample space.

One card can be drawn from a well-shuffled pack of 52 cards in $^{52}C_1$ ways i.e. 52 ways.

$\therefore$ n(S) = 52

Let E be the event of getting a club.

One club can be drawn in 13_{C_1} ways, i.e., 13 ways.

$\therefore$ n(E) = 13

$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{13}{52} = \frac{1}{4}$

Ans to Q No 2: Total number of balls = 6 + 8 + 10 = 24. One ball can be drawn in $^{24}C_1$ ways, i.e., 24 ways.

Let S be the sample space.

$\therefore$ n(S) = 24

(i) Let E_1 be the event of drawing a red ball.

One red ball can be drawn in 6_{C_1} ways, i.e., 6 ways.

$\therefore$ n(E_1) = 6

$\therefore P(E_1) = \frac{n(E_1)}{n(S)} = \frac{6}{24} = \frac{1}{4}$

(ii) Let E_2 be the event of drawing a black ball.

One black ball can be drawn in 8_{C_1} ways, i.e., 8 ways.

$\therefore$ n(E_2) = 8

$\therefore P(E_2) = \frac{n(E_2)}{n(S)} = \frac{8}{24} = \frac{1}{3}$

(iii) Let E_3 be the event of drawing a white ball.

There is no white ball in the bag.

$\therefore E_3 = f$

$\therefore P(E_3) = P(f) = 0$

Ans to Q No 3: Let S be the sample space.

$\therefore$ S = {HHH, HTH, HHT, THH, HTT, TTH, THT, TTT}

(i) Let E_1 be the event of getting 2 heads.

$\therefore$ E_1 = { HTH, HHT, THH }

$\therefore P(E_1) = \frac{n(E_1)}{n(S)} = \frac{3}{8}$

(ii) Let E_2 be the event of getting at least 2 heads.

$\therefore$ E_2 = { HTH, HHT, THH, HHH }

$$\therefore P(E_2) = \frac{n(E_2)}{n(S)} = \frac{4}{8} = \frac{1}{2}$$

Ans to Q No 4: Let S be the sample space.

One number can be chosen from the first 150 natural numbers in ${}^{150}C_1$ ways, i.e., 150 ways.

$\therefore$ n(S) = 150

S = { 1, 2, 3, 4, ..., 149, 150 }

Let E be the event of choosing a multiple of 5.

$\therefore$ E = { 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150 }

$\therefore$ n(E) = 30

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{30}{150} = \frac{1}{5}$$

Ans to Q No 5: Let S be the sample space

$\therefore$ n(S) = 11

Let E be the event of getting a vowel

$\therefore$ n(E) = 4

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{4}{11}$$

Ans to Q No 6: Let S be the sample space

$\therefore$ n(S) = 6 x 6 = 36

(i) Let E_1 be the event of getting a total of 6.

$\therefore$ E_1 = { (1, 5), (2, 4), (3, 3), (4, 2), (5, 1) }

$n(E_1) = 5$

$$\therefore P(E_1) = \frac{n(E_1)}{n(S)} = \frac{5}{36}$$

(ii) Let E_2 be the event of getting a total of 12.

$\therefore$ E_2 {(6, 6)}

$$\therefore P(E_2) = \frac{n(E_2)}{n(S)} = \frac{1}{36}$$

(iii) Let E_3 be the event of getting a total of 18.

Clearly $E_3 = f$, there cannot be total of 18.

$\therefore P(E_3) = P(f) = 0$

Ans to Q No 7: Two cards can be drawn in ${}^{52}C_2$ ways.

Let S be the sample space.

$$\therefore n(S) = {}^{52}C_2 = \frac{52!}{2!50!} = \frac{52\times51\times50!}{2\times1\times50!} = 26\times51$$

There are 4 kings, and 4 queens.

One king can be drawn in ${}^{4}C_1$, i.e., 4 ways.

One queen can be drawn in ${}^{4}C_1$, i.e., 4 ways.

Let E be the event of getting a king and a queen.

$\therefore$ n(E) = 4 x 4 = 16

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{16}{26\times51} = \frac{8}{13\times51} = \frac{8}{663}$$

Ans to Q No 8: Let S be the sample space

$$\therefore n(S) = {}^{52}C_2 = 26\times51$$

Let E be event of getting 2 hearts.

$$\therefore n(S) = {}^{13}C_2 = \frac{13!}{2!11!} = \frac{13\times12\times11!}{2\times1\times11!} = 13\times6$$

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{13\times6}{26\times51} = \frac{1}{17}$$

Ans to Q No 9: Total number of balls = 4 + 5 + 6 = 15

2 balls can be chosen from 15 balls in ${}^{15}C_1$ ways.

Let S be the sample space.

$$\therefore n(S) = {}^{15}C_2 = \frac{15!}{2!13!} = \frac{15\times14}{2\times1} = 15\times7 = 105$$

If none of the balls is black, then the two balls must be from white and green balls.

Total number of white and green balls = 4 + 6 =10

Let E be the event of getting 2 balls, none of which is black.

$$\therefore n(E) = {}^{10}C_2 = \frac{10!}{2!8!} = \frac{10\times9}{2\times1} = 45$$

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{45}{105} = \frac{3}{7}$$

Ans to Q No 10: Let S be the sample space.

$\therefore$ n(S) = 26

Let E be the event of getting a letter that precedes 'h' and is a vowel.

$\therefore$ E = { a, e }

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{2}{26} = \frac{1}{13}$$

Ans to Q No 11: A leap year contains 366 days

366 days = 52 weeks + 2 days.

52 weeks contain 52 Mondays. So the 53rd Monday will be from the remaining 2 days.

The remaining 2 days will be any one of the following, – Sun and Mon, Mon and Tues, Tues and Wed, Wed and Thurs, Thurs and Fri, Fri and Sat, Sat and Sun. Out of these 7 possible outcomes, two outcomes are favomable (Sun and Mon, Mon and Tues)

$\therefore$ the required probability $= \frac{2}{7}$

Check Your Progress 2

Ans to Q No 1: Let S be the sample space.

$\therefore$ n(S) = 150

Let E_1 be the event of getting a multiple of 4.

$\therefore$ n(E_1) = 37

Let E_2 be the event of getting a multiple of 5.

$\therefore$ n(E_2) = 30

$\therefore E_1 \cap E_2$ is the event of getting a multiple of 4 and of 5, i.e., of 20.

$\therefore n(E_1 \cap E_2) = 7$

Now, $P(E_1 \cup E_2)$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} - \frac{n(E_1 \cap E_2)}{n(S)}$$

$$= \frac{37}{150} + \frac{30}{150} - \frac{7}{150}$$

$$= \frac{60}{150} = \frac{2}{5}$$

Ans to Q No 2: Let S be the sample space

$\therefore n(S) = {}^{52}C_1 = 52$

Let E_1 be the event of getting a spade.

$\therefore n(E_1) = {}^{13}C_1 = 13$

Let E_2 be the event of getting a king.

$\therefore n(E_2) = {}^{4}C_1 = 4$

$E_1 \cap E_2$ is the event of getting a spade and a king.

$\therefore n(E_1 \cap E_2) = 1$

$\therefore$ P (getting a spade or a king)

$= P(E_1 \cup E_2)$

$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$

$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} - \frac{n(E_1 \cap E_2)}{n(S)}$

$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$

Ans to Q No 3: We have $P(E_1) = 0.3$, $P(E_2) = 0.2$, $P(E_1 \cap E_2) = 0.1$

(i) $P\left(\overline{E_1 \cap E_2}\right) = 1 - P(E_1 \cap E_2) = 1 - 0.1 = 0.9$

(ii) $P\left(\overline{E_1 \cup E_2}\right) = 1 - P(E_1 \cup E_2)$

$= 1 - [P(E_1) + P(E_2) - PE_1 \cap E_2]$

$= 1 - 0.3 - 0.2 + 0.1 = 0.6$

Ans to Q No 4: We have P(not E_1) = 0.45

or $P(\overline{E_1}) = 0.45$

or $1 - P(E_1) = 0.45$

$\therefore P(E_1) = 1 - 0.45 = 0.55$

E_1, E_2 are two mutually exclusive events.

$\therefore P(E_1 \text{ or } E_2) = P(E_1 \cup E_2) = P(E_1) + P(E_2)$

$\therefore P(E_1) + P(E_2) = 0.65$

or, $0.55 + P(E_2) = 0.65$

$\therefore P(E_2) = 0.65 - 0.55 = 0.1$

Ans to Q No 5: We have $P(E_1) = \frac{1}{2}$, $P(E_2) = \frac{1}{3}$

P (E_1 and E_2) = $P(E_1 \cap E_2) = \frac{1}{6}$

(i) P(E_1 or E_2) = $P(E_1 \cup E_2)$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

(ii) P(neither E_1 nor E_2)

= P($\overline{E}_1$ and $\overline{E}_2$)

$$= P(\overline{E}_1 \cap \overline{E}_2)$$

$$= P\left(\overline{E_1 \cup E_2}\right)$$

$$= 1 - P(E_1 \cup E_2)$$

$$= 1 - \frac{2}{3} = \frac{1}{3}$$

Ans to Q No 6: Let S be the sample space.

$$\therefore n(S) = {}^6C_1 \times {}^6C_1 = 6 \times 6 = 36$$

Let E_1 be the event of getting a sum that is divisible by 3.

$\therefore$ E_1 = { (1, 2), (1, 5), (2, 1), (2, 4), (3, 3), (3, 6), (4, 2), (4, 5), (5, 1), (5, 4), (6, 3), 6, 6) }

$\therefore$ n (E_1) = 12.

Let E_2 be the event of getting a sum that is divisible by 4.

$\therefore$ E_2 = {(1, 3), (2, 2), (2, 6), (3, 1), (3, 5), (4,4), (5, 3), (6, 2), (6, 6)}

$\therefore$ n(E_2) = 9

$E_1 \cap E_2$ is the event of getting a sum that is divisible by 3 and 4.

$\therefore E_1 \cap E_2 = \{(6, 6)\}, \quad \therefore n(E_1 \cap E_2) = 1$

P(neither E_1 nor E_2)

= P($\overline{E}_1$ and $\overline{E}_2$)

$$= P(\overline{E}_1 \cap \overline{E}_2)$$

$$= P\left(\overline{E_1 \cup E_2}\right)$$

$$= 1 - P(E_1 \cup E_2)$$

$$= 1 - [P(E_1) + P(E_2) - P(E_1 \cap E_2)]$$

$$= 1 - \left[\frac{12}{36} + \frac{9}{36} - \frac{1}{36}\right]$$

$$= 1 - \frac{5}{9}$$

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



10.10 MODEL QUESTIONS

Q 1: In a simultaneous toss of a coin and a die, what is the sample space?

Q 2: Find the probability of getting a tail in a throw of a coin.

Q 3: Three coins are tossed together, what is the probability of getting (i) one head, (ii) at least one head, (iii) two heads, (iv) at least two heads, (v) all heads.

Q 4: In a simultaneous throw of two dice, find the probability of getting a total of 9 or 11 ?

Q 5: What is the probability that an ordinary year (not leap year) has 53 Mondays ?

Q 6: An urn contains 15 balls numbering from 1 to 15 (no two balls are of the same number). What is the probability that a ball selected at random is a ball with number that is multiple of 3 or 4 ?

*** ***** ***

UNIT 11: CONDITIONAL PROBABILITY

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Conditional probability

11.4 Multiplication theorem of probability

11.5 Law of probability

11.6 Baye's theorem

11.7 Let Us Sum Up

11.8 Further Readings

11.9 Answers to check your progress

11.10 Model Questions

11.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- define conditional probability
- know multiplication theorem of probability
- define law of probability
- know Baye's theorem

11.2 INTRODUCTION

In unit 10, we discussed about probbility and some important theorems related to probability. In this unit, we will discuss the concept of conditional probability. We will also discuss dependent and independent events. Multiplication theorems for dependent and independent events will also be discussed in this unit. Finally, we will disuss the law of probability and Baye's theorem.

11.3 CONDITIONAL PROBABILITY

In unit 10, we have discussed the methods of finding the probability of events. Supose, we have two events from the same sample space in a

random experiment, does the occurrence of one of the events affect the probability of the other event? Let us see this situation with the following examples

1. Consider a random experiment of throwing a die.

Let S be the sample space and A be the event of of getting an odd numbers.

$\therefore\ S=\{1,2,3,4,5,6\}$ and $A=\{1,3,5\}$

$$P(A)=\frac{3}{6}=\frac{1}{2}$$

Suppose we consider another event B such that B be the event of getting the numbers greater than 3. Then $B=\{4,5,6\}$.

When we say that B has already occurred,the event of getting ' numbers is less than or equal to 3'is not possible. Now, the number of elements of of the sample space is reduced to 3 i.e $\{4,5,6\}$. Out of these three points, two points are even number $\{4,6\}$.

Thus, given that if the event B has already occurred, then $P(A)=\frac{2}{3}$. Hence, the probability of A given that B has already occurred is denoted by $P\left(\frac{A}{B}\right)$.

2. Consider the experiment of tossing three fair coins. The sample space of the experiment is $S=\{HHH,HHT,HTH,THH,HTT,THT,TTH,TTT\}$. Let E be the event 'at least two tails appear' and F be the event 'first coin shows head'.

Now, suppose we are given that the first coin shows head, i.e. F occurs, then we have to find the probability of occurrence of E. With the information of occurrence of F, we are sure that the cases in which first coin does not result into a head should not be considered while finding the probability of E. This information reduces our sample space from the set S to its subset F for the event E. In other words, the additional information really amounts to telling us that the situation may be considered as being that of a new random experiment for which the sample space consists of all those outcomes only which are favourable to the occurrence of the event F. Now, the sample point of F which is favourable to event E is HTT. Thus, Probability of E considering F as the sample space $=\frac{1}{4}$.

Probability of E given that the event F has occurred $=\frac{1}{4}$

This probability of the event E is called the *conditional probability of E given that F has already occurred*, and is denoted by P (E|F).

3. Consider the random experiment of drawing a card from a deck of 52 cards. Suppose we draw a queen in successive draws. In first draw the probability of drawing a queen is $\frac{4}{52}$. If the card is not replaced before the second draw, then the event of getting queen in the second draw depends on the first event. Now, there will be only 3 queens. The probability of getting a queen in second draw will also be $\frac{3}{52}$.

In the above examples,we observe that the occurrence of one event depends on another event which has already happend. Hence,an event E is said to be a dependent event if the occurnce of A depends on another event F which has already happend. It is denoted by $P(E \mid F)$.

The above examples leads to the definition of conditional probability as :

Let E and F be two events defined on a sample space S. Let $P(F) > 0$,then the conditional probability of E given that F has already occurred denoted by $P(E \mid F)$ and mathematically written as :

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}, P(F) > 0$$

Similarly, $P(F \mid E) = \frac{P(E \cap F)}{P(E)}, P(E) > 0$

Note : (1) $P(E \mid F)$ and $P(F \mid E)$ can also be expressed as follows :

$$P(E \mid F) = \frac{n(E \cap F)}{n(F)} \quad \text{and} \quad P(F \mid E) = \frac{n(E \cap F)}{n(E)}.$$

(2) $P(E \mid F)$ is read as probability of event E given event F.

Laws of Conditional probability :

(a) If E_1 and E_2 be any two events of a sample space S and A be any event of S such that $P(A) \neq 0$, then

$$P(E_1 \cup E_2 \mid A) = P(E_1 \mid A) + P(E_2 \mid A) - P(E_1 \cap E_2 \mid A)$$

We have $P(E_1 \cup E_2 \mid A) = \dfrac{P((E_1 \cup E_2) \cap A)}{P(A)}$

$$= \frac{P(E_1 \cap A) \cup P(E_2 \cap A)}{P(A)}$$

$$= \frac{P(E_1 \cap A) + P(E_2 \cap A) - P(E_1 \cap E_2 \cap A)}{P(A)}$$

$$= \frac{P(E_1 \cap A)}{P(A)} + \frac{P(E_2 \cap A)}{P(A)} - \frac{P(E_1 \cap E_2 \cap A)}{P(A)}$$

$$= P(E_1 \mid A) + P(E_2 \mid A) - P(E_1 \cap E_2 \mid A)$$

(b) $P(E' \mid F) = 1 - P(E \mid F)$

We have $P(S \mid F) = 1$

$$\Rightarrow P(E' \cup E \mid F) = 1$$

$$\Rightarrow P(E' \mid F) + P(E \mid F) = 1$$

$$\Rightarrow P(E' \mid F) = 1 - P(E \mid F)$$

Independent events

Events are said to be independent if the occurrence of one does not affect the others. In the experiment of tossing a fair coin, the occurrence of the event 'head' in the first toss is independent of the occurrence of the event ' head' in the second toss and third toss.

Example 1 : A card is chosen at random from an ordinary deck of 52 cards.

Let E be the event that the cards is an ace and F the event that the card is a club. Then E and F are independent.

Dependent events:

If the occurrence of one event influences the occurrence of the other, then the second event is said to be dependent on the first. In the above example, if we do not replace the first ball drawn, this will change the composition of balls in the bag while making the second draw and therefore the event of 'drawing a red ball' in the second will depend on event (first ball is red or white) occurring in first draw. Similarly, if a person draw a card from a full pack and does not replace it, the result of the draw made afterwards will be dependent on the first draw.

Example : Two fair dice are tossed. Let E be the event that the first die is a 3, F the event that the sum is 6 and G the event that sum is 7. Then E and F are dependent, but E and G are independent.

Note: If the events A and B are independent, that is the probability of occurrence of any one of them $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$.

11.4 MULTIPLICATION THEOREM ON PROBABILITIES

Now, we discuss multiplication theorem on probabilities for both independent and dependent events.

Multiplication theorem on probabilities for independent events :

If two events E and F are independent, the probability that both of them occur is equal to the product of their individual probabilities. i.e $P(E \cap F) = P(E)P(F)$.

Note: (1)The theorem can be extended to three or more independent events. If E, F,G……. be independent events, then

$$P(E \cap F \cap G) = P(E)P(F)P(G).$$

(2) If E and F are independent then the complements of E and F are also independent. i.e $P\left(\bar{E} \cap \bar{F}\right) = P\left(\bar{E}\right)P\left(\bar{F}\right)$

Multiplication theorem for dependent events: If E and F be two dependent events, i.e the occurrence of one event is affected by the occurrence of the other event, then the probability that both A and B will occur is

$$P(E \cap F) = P(E)P(F \mid E).$$

Note: In the case of three events E, F, G

$P(E \cap F \cap G) = P(E)P(F \mid E)P(G \mid E \cap F)$ i.e., the probability of occurrence of E, F and G is equal to the probability of E times the probability of F given that E has occurred, times the probability of G given that both E and F have occurred.

Ilustrative Examples :

Example1: Let E and F are two events,such that $P(E) = 0.8$, $P(F) = 0.6$, $P(E \cap F) = 0.5$. Find the value of (i) $P(E \cup F)$ (ii) $P(F \mid E)$ (iii) $P(E \mid F)$.

Solution : (i) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

$= 0.8 + 0.6 - 0.5$

$= 0.9$

(ii) $P(F \mid E) = \frac{P(E \cap F)}{P(E)} = \frac{0.5}{0.8} = \frac{5}{8}$

(iii) $P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{0.5}{0.6} = \frac{5}{6}$.

Example 2 : If $P(E) = 0.5, P(F) = 0.3$ and $P(E \cap F) = 0.15$. Find $P(E \mid F)$.

Solution : By definition of conditional probability

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

$$= \frac{0.15}{0.3}$$

$$= 0.5$$

Example 3 : If $P(E) = 0.9; P(F \mid E) = 0.8,$ find $P(E \cap F)$.

Solution : By definition of conditional probability

$$P(F \mid E) = \frac{P(F \cap E)}{P(E)}$$

$$\Rightarrow P(F \cap E) = P(E).P(F \mid E)$$

$$= 0.9 \times 0.8$$

$$= 0.72.$$

Example 4 : Find the chance of drawing 2 white balls in succession from a bag containing 5 red and 7 white balls, the balls drawn not being replaced.

Solution : Let E be the event that ball drawn is white in the first draw and F be the event that ball drawn is white in the second draw.

$$\therefore P(F \cap E) = P(E).P(F \mid E)$$

Here $P(E) = \frac{7}{12}, P(F \mid E) = \frac{6}{11}$

$$\therefore P(F \cap E) = \frac{7}{12}.\frac{6}{11} = \frac{7}{22}.$$

Example 5 : In an examination 30% of the students have failed in Mathematics, 20% of the students have failed in chemistry and 10 % have failed in both Mathematics and Chemistry. A student is selected at random

(i) What is the probability that the students has failed in Mathematics if it is known that he has failed in Chemistry ?

Solution : Let E be the event of failing in Mathematics.

$$\therefore P(E) = 30\% = 0.3$$

Let F be the event of failing in Chemistry

$$\therefore P(F) = 20\% = 0.2$$

Let $E \cap F$ be the event of failing in both subjects.

$$P(E \cap F) = 10\% = 0.1$$

The probability that the student has failed in Mathematics given that he has failed in Chemistry is

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{0.1}{0.2}$$

$$= 0.5 .$$

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

Solution : Probability of drawing a black ball in the first attempt is

$$P(E) = \frac{3}{5+3} = \frac{3}{8}$$

Probability of drawing the second black ball given that the first ball drawn is black

$$P(F \mid E) = \frac{2}{5+2} = \frac{2}{7}$$

$\therefore$ The probability that both balls drawn are black is given by

$$P(E \cap F) = P(A) \times P(F \mid E) = \frac{3}{8} \times \frac{2}{7} = \frac{3}{28} .$$



CHECK YOUR PROGRESS -1

Q 1: Two fair coins are tossed. Given that at least one of them is heads, what is the probability that both of them are heads?

Q 2: Suppose that a box contains 8 red balls and 4 white balls. We randomly draw 3 balls from the box without replacement. Find

the probability that all 3 are red.

Q 3: A die is rolled, If the outcome is an odd number. What is the probability that it is prime?

Q 4: A family has two children.Find the probability that both children are girls given that at least one of them is a girl.

11.5 TOTAL PROBABILITY

Theorem of total probability : If E_1 and E_2 are mutually exclusive and exhaustive events and $P(E_1) \neq 0$, $P(E_2) \neq 0$, then for any event A of S,

$$P(A) = P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)$$

Proof : Since E_1 and E_2 are mutually exclusive and exhaustive events, we have $E_1 \cup E_2 = S$ and $E_1 \cap E_2 = f$.

$\therefore A = A \cap S$

$= A \cap (E_1 \cup E_2)$

$= (A \cap E_1) \cup (A \cap E_2)$ .... (1)

Now, $(A \cap E_1) \cap (A \cap E_2) = A \cap (E_1 \cap E_2)$

$= A \cap f$

$= f$

$\therefore$ $A \cap E_1$ and $A \cap E_2$ are mutually exclusive and exhaustive events.

$\therefore$ From (1),we have

$P(A) = P(A \cap E_1) + P(A \cap E_2)$

$= P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)$ [multiplication theorem of probability]

If E_1, E_2, E_3 are mutually exclusive and exhaustive events and $P(E_1) \neq 0$, $P(E_2) \neq 0$, $P(E_3) \neq 0$ then for any event A of S,

$$P(A) = P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2) + P(E_3)P(A \mid E_3)$$

11.6 BAYE'S THEOREM

The concept of conditional probability discussed earlier takes into account information about the occurrence of one event to predict the probability of another event. This concept can be extended to revise probabilities based on new information and to determine the probability that

a particular effect was due to specific cause. The procedure for revising these probabilities is known as Bayes theorem. The Principle was given by Thomas Bayes in 1763. By this principle, assuming certain prior probabilities, the posteriori probabilities are obtained.

Bayes' Theorem or Rule

Statement : Let $E_1, E_2, E_3, \cdots, E_n$ be a set of n mutually exclusive and collectively exhaustive events and $P(E_1), P(E_2), P(E_3), \cdots, P(E_n)$ are their corresponding probabilities. If A is another event such that P(A) is not zero and the priori probabilities $P(A \mid E_i)$ i =1,2…, n are also known. Then

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2) + \cdots + P(E_n)P(A \mid E_n)}$$

i.e., $$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)}$$

Proof : Let S be the sample space of the random experiment.

The events $E_1, E_2, E_3, \cdots, E_n$ being exhaustive.

$$\therefore S = E_1 \cup E_2 \cup \cdots \cup E_n$$

$$A = A \cap S \qquad [A \subset S]$$

$$= A \cap (E_1 \cup E_2 \cup \cdots \cup E_n)$$

$$= (A \cap E_1) \cup (A \cap E_2) \cup \cdots (A \cap E_n) \quad \text{[Distributive law]}$$

$$\Rightarrow P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n)$$

$$= P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2) + \cdots + P(E_n)P(A \mid E_n)$$

$$= \sum_{i=1}^{n} P(E_i)P(A \mid E_i) \quad \text{.... (1)}$$

Now, $P(A \cap E_i) = P(A)P(E_i \mid A)$

$$\Rightarrow P(E_i \mid A) = \frac{P(A \cap E_i)}{P(A)}$$

$$= \frac{P(E_i)P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)} \qquad \text{[using (1)]}$$

Note : (1) The probabilities $P(E_1), P(E_2), \cdots, P(E_n)$ are called as the 'a prori probabilites'.

(2) The probabilities $P(E_1 \mid A), P(E_2 \mid A), \cdots, P(E_n \mid A)$ are called'posteriori probabilities'.

Example 1 : A purse contains 6 silver coins and 3 gold coins.Another purse contains 4 silver coins and 5 gold coins.A purse is selected at random and a coin is drawn from it.What is the probability that it is a silver coin.

Solution : Let Let E = Event of selecting first purse

F = Event of selecting the second purse

A= Event of getting a silver coin

$\therefore P(E) = \frac{1}{2}$ and $P(F) = \frac{1}{2}$

Also, $P(E \mid A) = \frac{6}{9} = \frac{2}{3}$ and $P(F \mid A) = \frac{4}{9}$

$\therefore$ Required probability $P(A) = P(E)P(A \mid E) + P(F)P(A \mid F)$

$$= \frac{1}{2} \times \frac{6}{9} + \frac{1}{2} \times \frac{4}{9}$$

$$= \frac{10}{18}$$

$$= \frac{5}{9}$$

Example 2 : A box contains two coins. One coin is heads on both sides and the other is heads on one side and tails on the other. One coin is selected from the box at random and the face of one side is observed. If the face is heads, what is the probability that the other side is heads ?

Solution : Let E = Event of selecting two headed coin

F = Event of selecting the other coin (one head and one tail)

A= Event of getting the second side also head.

Now, we have to find $P(E \mid A)$.

Here, $P(E) = P(F) = \frac{1}{2}$

$$P(A \mid E) = 1, P(A \mid F) = \frac{1}{2}$$

By Baye's theorem, we have

$$P(E \mid A) = \frac{P(E)P(A \mid E)}{P(E)P(A \mid E) + P(F)P(A \mid F)}$$

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

$$= \frac{2}{3}.$$

Example 3 : Assume that a factory has two machines. Past records show that machine I produces 20% of the items and machine II produces 80% of the items. Further, it shows that 6% of the items produced by machine I are defective and only 1 % of the items produced by machine II were defective. If a defective item is drawn at random,what is the probability that it was produced by machine I ?

Solution : Let E_1, E_2 be the events of drawing an item produced by machines I and II repectively. Let A denote the event of drawing a defective item.

We have $P(E_1)=\frac{20}{100}=\frac{1}{5}$, $P(E_2)=\frac{80}{100}=\frac{4}{5}$

Also, we have $P(A\mid E_1)=\frac{6}{100}$, $P(A\mid E_2)=\frac{1}{100}$

From Bayes' theorem, the required probability is

$$P(E_1\mid A)=\frac{P(E_1).P(A\mid E_1)}{P(E_1)P(A\mid E_1)+P(E_2)P(A\mid E_2)}$$

$$= \frac{\frac{1}{5}.\frac{6}{100}}{\frac{1}{5}.\frac{6}{100}+\frac{4}{5}.\frac{1}{100}} = \frac{6}{6+4} = 0.6.$$

Example4 : A bolt is manufactured by 3 machines A,B and C.A turns out twice as many items as B and machines B and C produce equal number of items. 2 % of bolts produced by A and B are defective and 4 % of bolts produced by C are defective.All bolts are put into one stock pile and 1 is chosen from this pile.What is the probability that it is defective ?

Solution : Let E_1, E_2, E_3 be the events of bolts manufactured by machine A,B,C repectively.

Here we have to apply total theorem on probability.

B and C produce equal no. of bolts and A produces two times that of B.

$$\therefore \quad P(E_1) = 50\% = \frac{50}{100} = 0.50$$

$$P(E_2) = 25\% = \frac{25}{100} = 0.25$$

$$P(E_3) = 25\% = \frac{25}{100} = 0.25$$

Let D be the event of getting defective bolt.

$$P(D \mid E_1) = \frac{2}{100} = 0.02$$

$$P(D \mid E_2) = \frac{2}{100} = 0.02$$

$$P(D \mid E_3) = \frac{4}{100} = 0.04$$

By total theorem on probability.

$$\begin{aligned} P(D) &= P(E_1)P(D \mid E_1) + P(E_2)P(D \mid E_2) + P(E_3)P(D \mid E_3) \\ &= 0.5 \times 0.02 + 0.25 \times 0.02 + 0.25 \times 0.04 \\ &= 0.0250 \end{aligned}$$

Example 5 : A bag A contains 2 white and 3 red balls and bag B contains 4 white and 5 red balls.One ball is drawn at random at random from one of the bags and is found to be red,find the probability that it is drawn from bag B.

Solution : Here given that the ball drawn is red and we have to find the probability of the ball coming from bag B.

Let E_1 be the event of selecting bag A.

$$\therefore \quad P(E_1) = \frac{1}{2}$$

Let E_2 be the event of selecting bag B.

$$\therefore \quad P(E_2) = \frac{1}{2}$$

Let D be the event of selecting red ball.

Probability of drawing red ball from bag A is

$$P(D \mid E_1) = \frac{3}{5}$$

Similarly, probability of drawing red ball from bag B is

$$P(D \mid E_2) = \frac{5}{9}$$

We have to find the red ball was drawn from bag B i.e., $P(E_2 \mid D)$

By Baye's Theorem

$$P(E_2 \mid D) = \frac{P(E_2)P(D \mid E_2)}{P(E_1)P(D \mid E_1) + P(E_2)P(D \mid E_2)}$$

$$= \frac{\frac{1}{2}.\frac{5}{9}}{\frac{1}{2}.\frac{3}{5} + \frac{1}{2}.\frac{5}{9}}$$

$$= \frac{\frac{5}{18}}{\frac{3}{10} + \frac{5}{18}}$$

$$= \frac{\frac{5}{18}}{\frac{27+25}{90}}$$

$$= \frac{\frac{5}{18}}{\frac{52}{90}}$$

$$= \frac{25}{52}$$

Example 6 : In a bolt factory machines A, B, C manufacture repectively 25%, 35%, 40% of the total. Of their output 5%, 4%, 2% are defective bolts. A bolt is drawn at random and found to be defective. What is the probability that it was manufactured by machine B ?

Solution : Let E_1, E_2, E_3 be the events of bolts manufactured by machine A, B, C repectively.

$$P(E_1) = 25\% = \frac{25}{100} = 0.25$$

$$P(E_2) = 35\% = \frac{35}{100} = 0.35$$

$$P(E_3) = 40\% = \frac{40}{100} = 0.40$$

Let D be the event of drawing a defective bolt.

$$\therefore P(D \mid E_1) = \frac{5}{100} = 0.05$$ [defective from 'A']

$$P(D \mid E_2) = \frac{4}{100} = 0.04$$ [defective from 'B']

$$P(D \mid E_3) = \frac{2}{100} = 0.02$$ [defective from 'C']

Now, we have to find the probability of defective bolt manufactured by machine B. i.e., $P(E_2 \mid D)$.

By Baye's Theorem :

$$P(E_2 \mid D) = \frac{P(E_2)P(D \mid E_2)}{P(E_1)P(D \mid E_1) + P(E_2)P(D \mid E_2) + P(E_3)P(D \mid E_3)}$$

$$= \frac{0.35 \times 0.04}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.4 \times 0.02}$$



CHECK YOUR PROGRESS -2

Q 1: Assume that a factory has two machines. Past records show that machine I produces 20% of the items and machine II produces 80% of the items. Further, it shows that 6% of the items produced by machine I are defective and 1% of the items produced by machine II are defective. If a defective item is drawn at random, what is the probability that it was produced by machine I ?

Q 2: Three machines A, B, C produce repectively 60%, 30%,10% of the total number of items of a factory. The percentage of respective outputs of these machines are repectively 2%, 3% and 4%. An item is selected at random and is found to be defective.Find the probability that the item was produced by machine C.

Q 3: A bag X contains 2 white and 3 red balls and a bag Y contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and is found to be red. Find the probability that it was drawn from bag Y.

Q 4: Urn A contains 3 red and 4 white balls and Urn B has 4 red and 5 white balls. One urn is selected at random and one ball is taken out of it. Thrown ball is red. What is the probability that it was taken from urn B ?

Q 5: A factory has three production lines I, II and III contributing 20%, 30 % and 50 %respectively, to its total output. The percentages of substandard items produced by lines I, II and III are, respectively, 15, 10 and 2. If an item chosen at random from the total output is found to be substandard, what is the probability that the item is from line I ?.



11.7 LET US SUM UP

- For two events E and F, the probability of **an** event **E** under the condition that F has already occumh is the **conditional probability P(E|F)** of **E** and is given by $P(E \mid F) = \frac{P(E \cap F)}{P(F)}$.
- Two events for which the occurrence of one has no influence on the occurrence or non-occurrence of the other are **independent** events, otherwise, they are **dependent.**
- If the events E and F are independent and $P(E) > 0, P(F) > 0$, then $P(E \cap F) = P(E)P(F)$ and $P(E \mid F) = P(E)$, $P(F \mid E) = P(F)$.



11.8 FURTHER READING

1) Rao. G. Shanker, Probability and Statistics, Universitis Press.

2) Statistics, R.S.N. Pillai and Bagavathi S. Chand and Company Ltd. New Delhi-110055.

.



11.9 ANSWERS TO CHECK YOUR PROGRESS

CHECK YOUR PROGRESS -1

Ans to Q No 1: Let E be the event that both of them are heads and F the event that at least one of them is heads.

Then, $P(E)=\frac{3}{4}, P(E\cap F)=\frac{1}{4}, P(F)=\frac{1}{4}$

$$\therefore P(E\mid F)=\frac{P(E\cap F)}{P(F)}$$

$$=\frac{\frac{1}{4}}{\frac{3}{4}}$$

$$=\frac{1}{3}.$$

By using the definition of conditional probability, we can easily get that

$$P(E\cap F)=P(F)P(E\mid F)=\frac{1}{4}\times\frac{1}{3}=\frac{1}{12}.$$

Ans to Q No 2: For i= 1;2;3, let E_i be the event that thei-th ball is red.

Then, $P(E_1\cap E_2\cap E_3)=P(E_1)P(E_2\mid E_1)P(E_3\mid E_1E_2)$

$$= \frac{8}{12}.\frac{7}{11}.\frac{6}{10} = \frac{14}{55}.$$

Ans to Q No 3: When a die is rolled, the sample is $S = \{1,2,3,4,5,6\}$.

Let $E =$ Event of getting an odd number

$= \{1,3,5\}$

F = Event of getting a prime number

$= \{2,3,5\}$

Then, $E \cap F = \{3,5\}$

$\therefore P(E) = \frac{3}{6} = \frac{1}{2}, P(F) = \frac{3}{6} = \frac{1}{2}$ and $P(E \cap F) = \frac{1}{3}$.

P(getting a prime already when an odd number)

= P(getting a prime number | getting an odd number)

$= P(E \mid F)$

$= P(E \mid F)$

$$= \frac{P(E \cap F)}{P(F)}$$

$$= \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Ans to Q No 4: The sample space S is given by

$S = \{(b,b),(b,g),(g,b),(g,g)\}$

Let E : Event that both children are girls

F : Event that at least one of the child is a girl.

Then $E = \{(g,g)\}$ and $F = \{(b,g),(g,b),(g,g)\}$

$E \cap F = \{(g,g)\}$

$\therefore P(E) = \frac{3}{4}, P(E \cap F) = \frac{1}{4}$

$\therefore$ Required probability is $P(E \mid F) = \frac{P(E \cap F)}{P(F)}$

$$= \frac{\frac{1}{4}}{\frac{3}{4}}$$

$$= \frac{1}{3}$$

CHECK YOU PROGRESS-2

Ans to Q No 1: Let E_1, E_2 be the events of drawing an item produced by machines I and II repectively.Let A denote the event of drawing a defective item.

We have $P(E_1) = \frac{20}{100} = \frac{1}{5}, P(E_2) = \frac{80}{100} = \frac{4}{5}$

Also, we have $P(A \mid E_1) = \frac{6}{100}, P(A \mid E_2) = \frac{1}{100}$

From Bayes's Theorem, the required probability is

$$P(E_1 \mid A) = \frac{P(E_1)P(A \mid E_1)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{\frac{1}{5}.\frac{6}{100}}{\frac{1}{5}.\frac{6}{100} + \frac{4}{5}.\frac{1}{100}} = \frac{6}{6+4} = 0.6.$$

Ans to Q No 2: Let E, F, G denote the events in which the items are selected by machines A, B, C repectively.

Then, we have $P(E) = \frac{60}{100} = 0.6, P(F) = \frac{30}{100} = 0.3, P(G) = \frac{10}{100} = 0.1$

Let D denote the event in which the item selected at random is a defective item.Then

$$P(D \mid E) = \frac{2}{100} = 0.02, P(D \mid F) = \frac{3}{100} = 0.03, P(D \mid G) = \frac{4}{100} = 0.04$$

Applying Bayes's theorem, we obtain

$$P(G \mid D) = \frac{P(G)P(D \mid G)}{P(E)P(D \mid E) + P(F)P(D \mid F) + P(G)P(D \mid G)}$$

$$= \frac{(0.1)(0.04)}{(0.6)(0.02) + (0.3)(0.03) + (0.1)(0.04)}$$

$$= \frac{0.004}{0.012 + 0.009 + 0.004}$$

$$= \frac{0.004}{0.025}$$

$$= 0.16.$$

Ans to Q No 3: Let E_1 : the ball is drawn from bag X

E_2 : the ball is drawn from bag Y

and A ; the ball is red.

Here $P(E_1)=P(E_2)=\frac{1}{2}$

Also, $P(A|E_1)$ = P(a red ball is drawn from bag X)

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

$P(A|E_2)$ = P(a red ball is drawn from bag Y)

$$=\frac{5}{9}$$

By Baye's Theorem ,we have

$$P(E_2|A)=\frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$=\frac{\frac{1}{2}\times\frac{5}{9}}{\frac{1}{2}\times\frac{3}{5}+\frac{1}{2}\times\frac{5}{9}}=\frac{25}{52}.$$

Ans to Q No 4: Required probability $=\frac{28}{45}$ [By Bayes's Theorem]

Ans to Q No 5: By Bayes formula

$$P(E_1|A)=\frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$=\frac{0.2\times0.15}{0.2\times0.15+0.3\times0.10+0.5\times0.02}$$

$$=0.069.$$



11.10 MODEL QUESTIONS

Q 1: E and F are two events for which $P(E)=\frac{1}{2}$, $P(F)=\frac{1}{3}$ and

$P(E \cap F) = \frac{1}{4}$.Find (i) $P(E \mid F)$,(ii) $P(E \cup F)$ and (iii) $P(E' \mid F')$.

Q 2: Three persons A, B, C supply raw mterials to a factory in the proportion $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$ and $5\%, 6\%, 8\%$ of the materials supplied by them repectively are found to be defective. A material supplied is selected at random from the total supply. What are the probabilities that material was supplied by A, or by B or by C ?

Q 3: One of three identical coins is unbiased while the remaining two are biased with probabilities of getting head in a toss are $\frac{1}{3}$ and $\frac{2}{3}$. One of the coins selected randomly shows head when tossed. Find the probability of the coin to be unbiased.

Q 4: Boys from the districts of Kamrup, Nagaon and Jorhat come to a college in Assam in the proportions 50%, 30% and 20% repectively. 80%, 60% and 40% of the boys comingrepectively from the districts wear Dhooti. A boy selected randomly from all the boys of the college is found to wear Dhooti. What is the probability that he comes kamrup?

Q 5: Three trains X, Y, Z carry repectively 25%, 35%, 40% of the total passengers. Of the passengers carried by the tains 5%, 4% and 3% respectively are ladies. If a pssengers chosen at random out of all the passengers of the trains found to be a lady. Prove that the probability that she was carried by X, or Y or Z are repectively $\frac{25}{69}, \frac{28}{69}, \frac{16}{69}$.

Q 6: Three are three groups of children. The first group consists of 3 girls and 1 boy, the second consists of 2 girls and 2 boys and the third consists of 1 girl and 3 boys. One child is picked up at random from each group. Find the probability of selecting 1 girl and 2 boys.

Q 7: A bag contains 5 white and 7 red balls. If a ball is drawn twice at random successively, find the probability that one ball is white and the other is red.

Q 8: In a running competition, the probabilities of the two competitiors of

winning the race is $\frac{1}{3}$ and $\frac{1}{5}$. What is the probability that both of them will not win ?

Q 9: One card is drawn from a well shuffled pack of 52 cards.Deduce the probability of drawing.

(i) a red card or an ace or both

(ii) a diamond or an ace or both.

(iii) king of diamond or king of heart or both.

Q 10: 4 men play with a pack of cards containing 52 cards. What is the probability that the 4 kings are held by only one of the players ?

*** ***** ***

UNIT 12 : RANDOM VARIABLES AND PROBABILITY DISTRIBUTION

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Random Variables

12.3.1 Discrete Random Variables

12.3.2 Continuous Random Variables

12.4 Probability Distribution of Random Variables

12.4.1 Mean and Variance of Random Variables.

12.5 Let Us Sum Up

12.6 Further Readings

12.7 Answers To Check Your Progress

12.8 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, you will able to learn

- random variables
- discrete random variables
- continuous random variables
- probability distribution of random variables.

12.2 INTRODUCTION

Let 3 coins be tossed. One may be interested to know the number of heads obtained. Similarly, Let a pair of dice be thrown. One may be interested to know the sum of points. In each of the cases a real number as associated with the outcome. A rule that assigns a real number to each outcome is called a random variable.

12.3 RANDOM VARIABLES

Let S be a sample space (Collection of outcomes) associated with a given random experiment.

Let $w \in S$ i.e w is an outcome.

A rule X that associates **each** $w \in S$ **to a unique** real number X(w) is called a random variable. Thus, a random variable is a real-valued function X which assigns a **unique** real number X(w) to **each** $w \in S$**.**

Let 3 coins be flipped.

Let S be the sample space associated with this random experiment. Then.

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Let us consider the function

X : $S \to IR$, where IR is the set of real numbers such that.

X(w) = number of heads in w, for all $w \in S$

$\therefore$ X (HHH) = 3, X(HHT) = 2, X(HTH) = 2, X(THH) = 2

X(HTT) = 1, X(THT) = 1, X(TTH) = 1, X(TTT) = 0

Here, X is a random variables

Clearly, range (X) = {0, 1, 2, 3}

12.3.1 Discrete Random Variable

A random variable which can take only finite or countable infinite number of values is called a discrete rancom variable.

The example given above is discrete random variable. In this case the random variable X takes four values, viz, 0, 1, 2, 3.

12.3.2 Continuours Random Variable

A random variable which can take any value between two given limits in called a continuous random variable.

Let the maximum life of electric bulbs be 2000 hours. The continuous random variable is

$X(w) = \{x : 0 \le x \le 2000\}$

12.4 PROBABILITY DISTRIBUTION OF RANDOM VARIABLES

If the values of a random variable together with the corresponding probabilities are given, then this description is called a probability distribution of the random variable.

Let a random variable X take the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively. Then the probability distribution of X is given by

X	x_1	x_2	–	– – – –	x_n
P(x)	p_1	p_2	–	– – – – –	p_n

Note : The above probability distribution of X is defined only when

(a) each $p_i \geq 0$

(b) $\sum_{i=1}^{n} p_i = 1$

Example 1: There coins are tossed simultaneously. Let X be the random variable such that.

X(w) = number of heads in w.

Write down the probability distribution of X.

Solution : Let S be the sample space.

Then

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

We have n(S) = 8

Here X = number of heads in tossing 3 coins.

Since the number of heads may be 0, 1, 2 or 3, therefore the possible values of X are 0, 1, 2, 3.

$\therefore$ P(X = 0) = P (getting no head)

= P(TTT) = $\frac{1}{8}$

P(X=1) = P (getting 1 head)

= P(HTT or THT or TTH) = $\frac{3}{8}$

P(X=2) = P(getting 2 heads)

= P(HHT or HTH or THH} = $\frac{3}{8}$

P(X=3) = P(getting 3 heads)

= P(HHH) = $\frac{1}{8}$

$\therefore$ the probability distribution is

X	0	1	2	3
P(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Note : Here each $P_i > 0$

Also $P_1 + P_2 + P_3 + P_4 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$

Example 2 : A die is tossed. Let X be the random variable defined by

$$X = \begin{cases} 1, \text{ if the outcome is an even number} \\ 0, \text{if the outcome is an odd number} \end{cases}$$

Write down the probability distribution of X.

Solution: Let S be the sample space.

$\therefore$ S = {1, 2, 3, 4, 5, 6}

Here X takes the values 0, 1.

P(X = 0) = P (getting an odd number)

$= \frac{3}{6} = \frac{1}{2}$

P(X = 1) = P (getting an even number)

$= \frac{3}{6} = \frac{1}{2}$

$\therefore$ the probability distribution is

X	0	1
P(X)	$\frac{1}{2}$	$\frac{1}{2}$



CHECK YOUR PROGRESS 1

Q 1: Given an example of a discrete random variable.

Q 2: Give an example of a continuous random variable.

Q 3: Two coins are tossed simultaneously. Let X be the random variable such that
X(w) = number of heads in w. Write down the probability distribution of X.

Q 4: A die is tossed twice. 'Getting an even number on a toss' is considered a success. Find the probability distribution of number of successes.

Q 5: A die is tossed twice. 'Getting a number greater than 4' is considered a success. Find the probability distribution of number of successes.

Q 6: Two cards are drawn simultaneously from a well-shuffled pack of 52 cards. Find the probability distribution of number of queens.

Q 7: Two cards are drawn without replacement from a well-shuffled pack of 52 cards. Find the probability distribution of the number of face cards.

Q 8: 20% of the bulbs produced by a factory are defective. Find the probability distribution of the number of defective bulbs in a sample of 4 bulbs chosen at random.

12.4.1 Mean and Variance of Random Variables

Let a random variable X take the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively such that each $p_i \geq 0$, and $\sum_{i=1}^{n} p_i = 1$

The mean of x, denoted by m, is defined by

$$m = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i} = \sum_{i=1}^{n} x_i p_i, \mathbf{Q} \sum_{i=1}^{n} p_i = 1$$

The mean of x is also called expected value of X. It is denoted by E(X). Thus,

$$m = E(X) = \sum_{i=1}^{n} x_i p_i$$

The variance of X, denoted by s^2, is given by $s^2 = \Sigma x_i^2 p_i - m^2$

The standard deviation of X, denoted by s, is the positive square root of variance.

Example 3: Find the mean and the standard defiation for the following probability distribution.

X	1	3	5
P(X)	0.3	0.4	0.3

Solution : Mean m = 1 x 0.3 + 2 x 0.4 + 5 x 0.3

= 0.3 + 1.2 + 1.5

$SD = s = \sqrt{\Sigma x_i^2 p_i - m^2}$

Now $\Sigma x_i^2 p_i = 1 \times 0.3 + 9 \times 0.4 + 25 \times 0.3$

= 0.3 + 3.6 + 7.5

= 11.4

$\therefore SD = s = \sqrt{11.4 - 9} = \sqrt{2.4}$

Example 4 : A random variable X has the following probability distribution

X	0	1	2	3
P(X)	a	3a	3a	a

Find the value of a. Also find the mean.

Solution : We know that $\Sigma p_i = 1$

$\therefore$ a + 3a + 3a + a = 1

or 8a = 1

$\therefore a = \frac{1}{8}$

2nd part : Mean $m = \Sigma x_i p_i$

$$= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8}$$

$$= \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = \frac{3}{2}$$

Example 5 : A random variable X has the following probability distribution.

X	0	1	2	3	4
P(X)	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Find the mean of X. Also find (i) P (X > 3), (ii) P(0 < X < 3)

Solution : Mean $m = 0\times\frac{1}{16}+1\times\frac{1}{4}+2\times\frac{3}{8}+3\times\frac{1}{4}+4\times\frac{1}{16}$

$$=\frac{1}{4}+\frac{3}{4}+\frac{3}{4}+\frac{1}{4}=2$$

(i) $P(X>3)=P(X=4)=\frac{1}{16}$

(ii) $P(0<X<3)=P(X=1)+P(X=2)$

$$=\frac{1}{4}+\frac{3}{8}=\frac{2+3}{8}=\frac{5}{8}$$



CHECK YOUR PROGRESS 2

Q 1: A random variable X has the following probability distribution.

X	0	1	2	3
P(X)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

Find the mean and the standard deviation.

Q 2: Two cards are drawn simultaneously from a well-shuffled pack of 52 cards. Find the mean and the standard deviation of the number of aces.



12.5 LET US SUM UP

- Let S be a sample space associated with a given random experiment. A rule X that associates each $w \in S$ to a unique real number X(w) is called a random variable.
- A random variable which can take only finite or countably infinite number of values is called a discrete random variable.
- A random variable which can take any value between two given limits

is called a continuous random variable.

- If the values of a random variable together with the corresponding probabilities are given, then this description is called a probability distribution of the random variable.
- Let a random variable X take the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively. The mean of x, denoted by m, is defined by $m = \sum_{i=1}^{n} x_i p_i$

  The variance of X, denoted by s^2, is given by $s^2 = \Sigma x_i^2 p_i - m^2$

  The standard deviation of X, denoted by s, is the positive square root of variance.



12.6 FURTHER READING

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009
2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011
3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005
4) Probability, Random Process, Queueing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042
5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33



12.7 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: An example of a discrete random variable :

Three coins are tossed simultaneously. Let X be the the random variable such that X(w) = number of heads in w.

In this case the random variable X takes from values, viz, 0, 1, 2, 3.

So, it is a discrete random variable.

Ans to Q No 2: An example of a continuous random variable :

Let the maximum life of electric bulbs be 2000 hours. The continuous random variable is $x(w) = \{x : 0 \le x \le 2000\}$

Ans to Q No 3: Let S be the sample space.

$\therefore$ S = {HH, HT, TH, TT}

We have n(S) = 4

Here X = number of heads in tossing 2 coins.

Since the number of heads may be 0, 1, or 2, therefore the possible values of X are 0, 1, 2.

$\therefore$ P(X = 0) = P (getting no head)

$$= P(TT) = \frac{1}{4}$$

P (x = 1) = P (getting 1 head)

$$= P(HT \text{ or } TH) = \frac{2}{4} = \frac{1}{2}$$

P (x = 2) = P (getting 2 heads)

$$= P(HH) = \frac{1}{4}$$

$\therefore$ the probability distribution is

X	0	1	2
P(X)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Ans to Q No 4: Let the random variable, i.e., the number of successes be X.

On flipping a die twice, the number of successes can be 0, 1, or 2.

Thus, X can take values 0, 1, 2.

P(X = 0) = P (no success in both flips)

= P (no success in 1st flip). P(no success in 2nd flip)

$$= \frac{3}{6} \cdot \frac{3}{6} = \frac{1}{4}$$

P(X = 1) = P(success in 1st toss and no-success in 2nd toss) or P(no-success in 1st toss and success in 2nd toss)

$$=\frac{3}{6}\cdot\frac{3}{6}+\frac{3}{6}\cdot\frac{3}{6}=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

P(X = 2) = P (success in both the flips)

=P(success in 1st flip). P(success in 2nd flip)

$$=\frac{3}{6}\cdot\frac{3}{6}=\frac{1}{4}$$

$\therefore$ the probability distribution is

X	0	1	2
P(X)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Ans to Q No 5: Let the random variable, i.e., the number of successes be X.

One flipping a die twice, the number of successes can be 0, 1, or 2.

$\therefore$ X can take values 0, 1, 2.

P(X = 0) = P (no success in both flips)

= P(no success in 1st flip). P(no success in 2nd flip)

= P(getting a number ≤ 4). P (getting a number ≤ 4) $=\frac{4}{s}\cdot\frac{4}{s}=\frac{4}{9}$

P(X = 1) = P(success in 1st flip and no-success in 2nd slip) or P (no-success in 1st flip and success in 2nd slip)

= P(getting a number ≤ 4 and getting a number > 4) + P(getting a number > 4, and getting a number ≤ 4)

$$=\frac{4}{6}\cdot\frac{2}{6}+\frac{2}{6}\cdot\frac{4}{6}=\frac{4}{9}$$

P(X = 2) = P (success in both the flips)

= P(success in 1st flip). P(success in 2nd flip)

$$=\frac{2}{6}\cdot\frac{2}{6}=\frac{1}{9}$$

$\therefore$ the probability distribution is

X	0	1	2
P(X)	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

Ans to Q No 6: [There are 4 queens in a pack of 52 cards.]

Let the random variable, i.e., the number of queens be X.

On drawing 2 cards, the number of queens can be 0, 1, or 2.

∴ X can take values 0, 1, 2.

P(X = 0) = P(getting no queen)

= P(2 non-queens out of 48)

$$= \frac{^{48}C_2}{^{52}C_2} = \frac{188}{221}$$

P(X = 1) = P(getting 1 queen and 1 non-queen)

$$= \frac{^{4}C_1 . ^{48}C_1}{^{52}C_2} = \frac{32}{221}$$

P(X = 2) = P(getting 2 queens)

$$= \frac{^{4}C_2}{^{52}C_2} = \frac{1}{221}$$

∴ the probability distribution is

X	0	1	2
P(X)	$\frac{188}{221}$	$\frac{32}{221}$	$\frac{1}{221}$

Ans to Q No 7: [The jack, queen, and the king are the face cards. There are 4 jacks, 4 queens, and 4 kings. Thus, there are altogether 12 face cards.]

Let the random variable, i.e., the number of faces be X.

On drawing 2 cards, the number of faces can be 0, 1, or 2.

∴ X can take values 0, 1, 2.

P(X = 0) = P (getting no face cards)

= P (2 non-face cards out of 40)

$$= \frac{^{40}C_2}{^{52}C_2} = \frac{10}{17}$$

P(X = 1) = P(getting 1 face card and 1 non-face card)

$$= \frac{^{12}C_1 . ^{40}C_1}{^{52}C_2} = \frac{80}{221}$$

P(X = 2) = P (getting 2 face cards)

$$= \frac{^{12}C_2}{^{52}C_2} = \frac{11}{221}$$

∴ the probability distribution is

X	0	1	2
P(X)	$\frac{10}{17}$	$\frac{80}{221}$	$\frac{11}{221}$

Ans to Q No 8: Let the number of bulbs be 100.

∴ number of defective bulbs is 20, and the number of non-defective bulbs is 80.

Let the random variable, i.e., the number of defective bulbs be X.

On choosing 4 bulbs, the number of defective bulbs can be 0, 1, 2, 3 or 4.

∴ X can take values 0, 1, 2, 3, 4.

P(X = 0) = P(Choosing no defective bulbs)

$$= \frac{^{80}C_4}{^{100}C_4}$$

P(X = 1) = P(Choosing 1 defective bulb and 3 non-defective bulbs)

$$= \frac{^{20}C_1 \times ^{80}C_1}{^{100}C_4}$$

P(X = 2) = P(Choosing 2 defective bulbs and 2 non-defective bulbs)

$$= \frac{^{20}C_2 \times ^{80}C_2}{^{100}C_4}$$

P(X = 3) = P(Choosing 1 defective bulb and 3 non-defective bulbs)

$$= \frac{^{20}C_3 \times ^{80}C_1}{^{100}C_4}$$

P(X = 4) = P(Choosing 4 defective bulbs)

$$= \frac{^{20}C_4}{^{100}C_4}$$

∴ the probability distribution is

X	0	1	2	3	4
P(X)	$\frac{^{80}C_4}{^{100}C_4}$	$\frac{^{20}C_1 \times ^{80}C_3}{^{100}C_4}$	$\frac{^{20}C_2 \times ^{80}C_2}{^{100}C_4}$	$\frac{^{20}C_3 \times ^{80}C_2}{^{100}C_4}$	$\frac{^{20}C_4}{^{100}C_4}$

Check Your Progress 2

Ans to Q No 1: The probability distribution is

X	0	1	2	3
P(X)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{10}$

Mean $m = \Sigma x_i p_i$

$$= O\times\frac{1}{6}+1\times\frac{1}{2}+2\times\frac{3}{10}+3\times\frac{1}{30}$$

$$=\frac{1}{2}+\frac{3}{5}+\frac{1}{10}=\frac{5+6+1}{10}$$

$$=\frac{12}{10}=1.2$$

$$SD = s = \sqrt{\Sigma x_i^2 p_i - m^2}$$

Now $\Sigma x_i^2 p_i$

$$=0\times\frac{1}{6}+1\times\frac{1}{2}+4\times\frac{3}{10}+9\times\frac{1}{30}$$

$$=\frac{1}{2}+\frac{6}{5}+\frac{3}{10}=\frac{5+12+3}{10}=\frac{18}{10}=1.8$$

$$\therefore\ SD = a = \sqrt{1.8-1.44}=\sqrt{0.36}=0.6$$

Ans to Q No 2: [There are 4 aces in a pack of 52 cards]

Let the random variable, i.e., the number of aces be X.

On drawing 2 cards, the number of aces can be 0, 1, or 2.

$\therefore$ X can take values 0, 1, 2

P(X = 0) = P(getting no ace)

= P (2 non-aces out of 48)

$$=\frac{^{48}C_4}{^{52}C_2}=\frac{188}{221}$$

P(X = 1) = P(getting 1 ace and 1 non-ace)

$$=\frac{^{4}C_1\cdot{}^{48}C_1}{^{52}C_2}=\frac{32}{221}$$

P(X = 2) = P(getting 2 aces)

$$=\frac{^{4}C_1}{^{52}C_2}=\frac{1}{221}$$

$\therefore$ the probability distribution is

X	0	1	2
P(X)	$\frac{188}{221}$	$\frac{32}{221}$	$\frac{1}{221}$

Mean $m = \Sigma x_i p_i = 0 \times \frac{188}{221} + 1 \times \frac{32}{221} + 2 \times \frac{1}{221}$

$$= \frac{32}{221} + \frac{2}{221} = \frac{34}{221} = \frac{2}{13}$$

$$SD = s = \sqrt{\Sigma x_i p_i - m^2}$$

Now $\Sigma x_i^2 p_i$

$$= 0 \times \frac{188}{221} + 1 \times \frac{32}{221} + 4 \times \frac{1}{221} = \frac{32}{221} + \frac{4}{221}$$

$$= \frac{36}{221}$$

$$\therefore SD = s = \sqrt{\frac{36}{221} - \frac{4}{169}} = \sqrt{\frac{400}{2873}}$$



12.8 MODEL QUESTIONS

Q 1: A die is thrown. Let X be the number of points thrown. Write down the probability distribution and find the mean.

Q 2: Find the mean, and the standard deviation of the number of heads in two tosses of a coin.

Q 3: Find the mean, and the standard deviation of the number of heads in three tosses of a coin.

Q 4: A die is tossed twice. 'Getting a number greater than 3' is considered a success. Find the probability distribution of number of successes. Also find the mean and the standard deviation of number of successes.

Q 5: Find the probability distribution of the number of doublets in four throws of a pair of dice.

[Meaning of the word ‘doublet’ in this context is as follows.

‘Either of two dice, each of which, when thrown, has the same number of spots on the face lying uppermost.’]

*** ***** ***

UNIT 13 : THEORETICAL DISTRIBUTION I

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Bernoulli Trial

13.4 Binomial (and Bernoulli) DIstribution

13.5 Poisson DIstribution

13.6 Let Us Sum Up

13.7 Further Readings

13.8 Answers To Check Your Progress

13.9 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will able to-

- bernoulli trial
- binomial (and bernoulli) distribution
- possion distribution.

13.2 INTRODUCTION

In this unit, we will study theoretical discrete univariate (of only one variable) distribution. It gives us a law according to which different values of the random variable are distributed with specified probabilities. We will discuss two main distributions, viz., Binomial distribution, and poisson distribution.

13.3 BERNOULLI TRIALS

A Bernoulli trial (named after Jacob Bernoulli 1654-1705) is an experiment satisfying the following conditions.

1. There are only two possible outcomes, viz., success or failure (occurrence of an event is ‘success’. non-occurrence is ‘failure’.)
2. The prabability of a success (or failure) remains same in each trial.

3. Trails must be independent.
4. Number of trails must be finite.
5. The sum of probabilities of success and failure is 1.
6. The probability of succes vis-a-vis failure is not very low.

13.4 BINOMIAL (AND BERNOULLI) DISTRIBUTION

Let there be n Bernoulli trials. Let the random variable X denote the number of successes in these trials. Let the probability of getting a success in a single trial be p, and that of getting a failure be q so that p + q = 1. Then $P(X=r) = {}^nC_r\, p^r\, q^{n-r}$

Proof : Let a success be denoted by S, and a failure be denoted by F.

Number of ways of getting r successes in n trials = ${}^nC_r$

$\therefore P(X = r) = {}^nC_r$. P(SSS . . . . . S and FFF . . . . . F)

r times (n - r) times)

$= {}^nC_r$. P (S) P (S) . . . . P(S) . P(F) P(F) . . . . P(F)

r times (n - r) times)

$= {}^nC_r\, p^r\, q^{n-r}$ (P(S) = p, P(F) = q)

Note :

1. $P(x = 0) = {}^nC_0\, p^0\, q^{n-0} = q^n$

   $P(x = 1) = {}^nC_1\, p\, q^{n-1}$

   $P(x = 2) = {}^nC_2\, p^2\, q^{n-2}$

   - - - - - - - - - - - -

   - - - - - - - - - - - -

   - - - - - - - - - - - -

   $P(x = n) = {}^nC_n\, p^n\, q^{n-n} = p^n$

   $\therefore$ the distribution can be expressed as

X	0	1	2	. . . . .	n
P(X)	q^n	${}^nC_1\, pq^{n-1}$	${}^nC_2\, p^2q^{n-2}$	. . . . .	p^n

2. The distribution is called Binomial distribution since the probabilities of 0, 1, 2, ..., n successes, viz. q^n, ${}^nC_1\, pq^{n-1}$, ${}^nC_2\, p^2q^{n-2}$, ..., p^n, are the successive terms of the binomial expression $(q + p)^n$.
3. Sum of the probabilities

$= q^n + {}^nC_1\, pq^{n-1} + {}^nC_2\, p^2 q^{n-2} + \ldots . + p^n$

$= (q + p)^n = 1$

4. X can take only the integral values, viz,. 0, 1, 2, ... , n. Therefore, Binomial distribution is a discrete distribution.
5. Any random variable which follows Binomial distribution is called a binomial variate.
6. We have

   $P(X = r) = {}^nC_r\, p^r\, q^{n-r}$

   Thus, P(X = r) depends only on n, and p(or q). These constants are called parameters. Thus, Binomial distribution has two parameters.

Remark : The Bernoulli distribution is a special case of the binomial distribution where n = 1.

Example 1 : Find the probability distribution of the number of heads in 6 tosses of a coin.

Solution : If a coin is tossed, S = { H, T}

$\therefore$ P (getting a head) $= \frac{1}{2}$,

P (not getting a head) $= 1 - \frac{1}{2} = \frac{1}{2}$

Let X be the random variable denoting the number of heads.

In 6 trials, there may be 0 or 1 or 2 or 3 or 4 or 5 or 6 heads.

$$P(X = 0) = {}^6C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{6-0} = \frac{1}{64}$$

$$P(X = 1) = {}^6C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{6-1} = \frac{3}{32}$$

$$P(X = 2) = {}^6C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2} = \frac{15}{64}$$

$$P(X = 3) = {}^6C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{6-3} = \frac{5}{16}$$

$$P(X = 4) = {}^6C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{6-4} = \frac{15}{64}$$

$$P(X = 5) = {}^6C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{6-5} = \frac{3}{32}$$

$$P(X=6) = {}^6C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^{6-6} = \frac{1}{64}$$

$\therefore$ the probability distribution is

X	0	1	2	3	4	5	6
P(X)	$\frac{1}{64}$	$\frac{3}{32}$	$\frac{15}{64}$	$\frac{5}{16}$	$\frac{15}{64}$	$\frac{3}{32}$	$\frac{1}{64}$

Note : Sum of the probabilities

$$= \frac{1}{64} + \frac{3}{32} + \frac{15}{64} + \frac{5}{16} + \frac{15}{64} + \frac{3}{32} + \frac{1}{64} = 1$$

Example 2 : A die is thrown 3 times. Find the probability of getting at least one six.

Solution : If a die is thrown, S = { 1, 2, 3, 4, 5, 6}

$\therefore$ P (getting a six) $= \frac{1}{6}$

P (not getting a six) $= 1 - \frac{1}{6} = \frac{5}{6}$

Let X be the random variable denoting the number of sixes.

In 3 trials, there may be 0, or 1, or 2, or 3 sixes.

$\therefore$ the probability of getting at least 1 six

= P(X = 1) + P(X = 2) + P(X = 3)

$$= {}^3C_1\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{3-1} + {}^3C_2\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{3-2} + {}^3C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{3-3}$$

$$= \left(3 \times \frac{1}{6} \times \frac{25}{36}\right) + \left(3 \times \frac{1}{36} \times \frac{5}{6}\right) + \left(1 \times \frac{1}{216} \times 1\right)$$

$$= \frac{25}{72} + \frac{5}{72} + \frac{1}{216} = \frac{75+15+1}{216} = \frac{91}{216}$$

Mean of a Binomial Distribution :

Let a random variable X take the values x_1, x_2, x_3, ..., x_n with probabilities p_1, p_2, ..., p_n respectively. We know that the mean m is defined by

$$m = \sum_{i=1}^{n} x_i p_i$$

Now, let us consider the binomial distribution.

$P(X = r) = {}^nC_r\, p^r q^{n-r}, \quad r = 0, 1, 2, 3, \ldots, n.$

$$\therefore m = \sum_{r=0}^{n} r \,.\, P(X = r)$$

$$= \sum_{r=0}^{n} r \,.\, {}^nC_r\, p^r q^{n-r}$$

$$= 0 + 1.{}^nC_1 pq^{n-1} + 2^nC_2 p^2 q^{n-2} + 3^nC_3 p^3 q^{n-3} + \ldots\ldots + {}^n n.\,{}^nC_n p^n q^{n-n}$$

$$= 1.np.q^{n-1} + n(n-1)p^2 q^{n-2} + \frac{n(n-1)(n-2)}{2!} p^3 q^{n-3} + \ldots\ldots + np^n$$

$$= np\left[1.q^{n-1} + (n-1)pq^{n-2} + \frac{n(n-1)(n-2)}{2!} p^2 q^{n-3} + \ldots\ldots + p^{n-1}\right]$$

$$= np\left[{}^{n-1}C_0\, q^{n-1} + {}^{n-1}C_1\, q^{(n-1)-1}\; p + {}^{n-1}C_2\, q^{(n-1)-2} p2 + \ldots\ldots + {}^{n-1}C_{n-1}\, p^{n-1}\right]$$

$$= np\,(q + p)^{n-1}$$

$$= np, \quad \therefore q + p = 1$$

Variance of a Binomial Distribution

We know that the variance s^2 is defined by

$$s^2 = \sum_{i=1}^{n} r_i^2 p^i - m^2$$

So for the above binomial distribution,

$$\text{Variance} = s^2 = \sum_{r=0}^{n} r^2 p(X = r) - m^2$$

$$= \sum_{r=0}^{n} r^2\, {}^nC_r\, p^r q^{n-r} - n^2 p^2$$

$$= \sum_{r=0}^{n} \left[r + r(r-1)\right] {}^nC_r\, p^r q^{n-r} - n^2 p^2$$

$$= \sum_{r=0}^{n} r \,.\, {}^nC_r\, p^r q^{n-r} + \sum_{r=0}^{n} r(r-1)^n C_r\, p^r q^{n-r} - n^2 p^2$$

$$= np + \sum_{r=2}^{n} r(r-1)\frac{n(n-1)}{r(r-1)}\, {}^{n-2}C_{r-2} p^2 p^{r-2} q^{n-r} - n^2 p^2$$

$$= np + n(n-1)p^2 \sum_{r=2}^{n} {}^{n-2}C_{r-2}\, p^{r-2} q^{n-r} - n^2 p^2$$

$$= np + n(n-1)p^2 - n^2 p^2$$

$$= np + n^2 p^2 - np^2 - n^2 p^2$$

$$= np - np^2$$

$$= np(1 - p)$$

$= npr, \quad \mathbf{Q}\ p+q=1$

Thus, variance $= s^2 = npq$

$\therefore\ SD = s = \sqrt{npq}$

Recurrence relation for a binomial distribution :

We know that

$P(X=r) = {}^nC_r\, p^r q^{n-r}$

$\therefore\ P(X=r+1) = {}^nC_{r+1} p^{r+1} q^{n-r-1}$

$$\therefore \frac{P(X=r+1)}{P(X=r)} = \frac{{}^nC_{r+1} p^{r+1} q^{n-r-1}}{{}^nC_r\, p^r q^{n-r}}$$

$$= \frac{n!}{(r+1)!\,(n-r-1)!} \frac{r!(n-r)!}{n!} \cdot \frac{p}{q}$$

$$= \frac{n-r}{r+1} \cdot \frac{p}{q}$$

$$\therefore\ P(X=r+1) = \frac{(n-r}{(r+1)} \cdot \frac{p}{q} \cdot P(X=r)$$

Example 3 : The mean of a binomial distribution is 5, and the variance is 2.5. Find the distribution.

Solution : Let X be a binomial distribution with parameters n, and p.

We know that the mean is np, and the variance is npq, q = 1 – p.

$\therefore$ np = 5, and npq = 2.5

$\therefore$ 5q = 2.5 $\quad \Rightarrow q = \frac{2.5}{5} = \frac{1}{2}$

$\therefore$ p = 1 – q = 1 – $\frac{1}{2} = \frac{1}{2}$

$\therefore$ n = 10

$\therefore$ the distribution is

$${}^{10}C_r \left(\frac{1}{2}\right)^r \cdot \left(\frac{1}{2}\right)^{10-r}, \qquad 0 \le r \le 10\,.$$

Example 4 : The mean of a binomial distribution is 4 and the variance is $\frac{4}{3}$. Find $P(X \ge 1)$.

Solution : We have np = 4, and npq = $\frac{4}{3}$

$$\therefore\ 4q = \frac{4}{3} \Rightarrow q = \frac{1}{3}$$

$$\therefore\ p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$\therefore\ \frac{2n}{3} = 4 \quad \Rightarrow n = 6$$

Now $P(X \geq 1) = 1 - P(X = 0)$

$$= 1 - {}^{6}C_0\, p^0 q^{6-0} = 1 - q^6 = 1 - \left(\frac{1}{3}\right)^6$$

$$= 1 - \frac{1}{729} = \frac{728}{729}$$

Property : For a binomial distribution, mean is greater than variance.

Proof : Let X be a binomial distribution with parameters n, and p.

Now, mean = np

$> npq,$ **Q** $0 < q < 1$

= variance.

Example 5 : Can there be a binomial distribution with mean 5, and variance 7.5 ?

Solution : We know that in a binomial distribution, mean > variance.

Now, 5 < 7.5

$\therefore$ there cannot be binomial distribution with mean 5, and variance 7.5.



CHECK YOUR PROGRESS 1

Q 1: Find the probability distribution of the number of tails in 4 tosses of a coin.

Q 2: Find the probability distribution of the number of doublets in three throws of a pair of dice.

Q 3: A coin is flipped 8 times. Find the probability of getting at least 2 heads.

Q 4: Find the binomial distribution whose mean is 3, and the variance is 2.

Q 5: Find P for a binomial random variable X, if n = 6, and 9p(X = 4) = P(X = 2). Also, find the mean and the variance of the distribution.

Q 6: Let X be a binomial variate such that the mean is double the standard deviation. If P(X = 0) = $\frac{1}{16}$, find the probability that X is at most 1.

Q 7: A coin is flipped 8 times. Find the probability of getting at least 7 heads.

Q 8: For a binomial distribution, the mean is 6 and the standard deviation is $\sqrt{2}$. Write down the distribution.

Q 9: The mean and variance of a binomial variate X with parameters n and p are 16 and 8. Find (i) P (X = 0), (ii) P (X = 1), (iii) $P(X \geq 2)$

Q 10: Are the following statements true ? Justify your answer.

(i) The mean of a binomial distribution is 6, and its standard deviation is 3.

(ii) The mean of a binomial distribution is 5, and its variance is 2.

13.5 POISSON DISTRIBUTION

(After the name of French mathematician simeon Devis Poisson, 1781-1840)

Poisson distribution is a limiting case of the binomial distribution under the following three conditions.

1. The number of trials n is indefinitely large, i.e., $n \to \infty$
2. The constant probability p of success for each trial is indefinitely small, i.e., $p \to 0$
3. np = m (say) is finite.

We have

$$P(r) = P(X = r) = {}^nC_r\, p^r q^{n-r}$$

We know that P(0) = q^n

$$\therefore \lim_{n\to\infty} p(0) = \lim_{n\to\infty} q^n$$

$$= \lim_{n\to\infty} (1-p)^n$$

$$= \lim_{n\to\infty} \left(1-\frac{m}{n}\right)^n, \qquad \text{from np = m}$$

$$= e^{-m}$$

Again, we know that $\frac{P(r+1)}{P(r)} = \frac{n-r}{r+1}.\frac{p}{q}$

$$\therefore \lim_{n\to\infty} \frac{P(r+1)}{P(r)} = \lim_{n\to\infty} \frac{n-r}{r+1}.\frac{p}{q}$$

$$= \lim_{n\to\infty} \frac{n-r}{r+1}.\frac{\frac{m}{n}}{1-\frac{m}{n}}$$

$$= \lim_{n\to\infty} \frac{1-\frac{r}{n}}{1-\frac{m}{n}}.\frac{m}{r+1} = \frac{m}{r+1}$$

Now, $\lim_{n\to\infty} P(r) = \lim_{n\to\infty} \frac{P(r)}{P(r-1)} . \lim_{n\to\infty} \frac{P(r-1)}{P(r-2)} \ldots\ldots \lim_{n\to\infty} \frac{P(1)}{P(0)} . \lim_{n\to\infty} P(0)$

$$= \frac{m}{r} . \frac{m}{r-1} \ldots\ldots \frac{m}{1} e^{-m}$$

$$= \frac{m^r e^{-m}}{r!}$$

Definition : A random variable X is said to follow a poisson distribution if it takes non-negative values with the probability given by

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

Note :

1. X can take only the integral values, viz., 0, 1, 2, ... Therefore, poisson distribution is a discrete distribution.
2. Poisson distribution has only one parameter, viz, m.

Some examples where poisson distribution may be applied :

1. Number of defective blades in a packet of 100 blades.
2. Number of printing mistakes at each page of a book.
3. Number of deaths from heart attack or snake bite.
4. Number of air accidents in India in one year.
5. Number of cars passing through a street-crossing during the busy hours of a day.

Mean of a poisson distribution :

Mean = m

$$= \sum_{x=0}^{\infty} xP(X=x)$$

$$= \sum_{x=0}^{\infty} x\frac{e^{-m}\ m^x}{x!}$$

$$= \sum_{x=0}^{\infty} x\frac{e^{-m}\ mm^{x-1}}{x(x-1)!}$$

$$= me^{-m}\sum_{x=1}^{\infty}\frac{m^{x-1}}{(x-1)!}$$

$$= me^{-m}\left[1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+....\infty\right]$$

$$= me^{-m}e^{m}$$

$$= m$$

Variance of a poisson distribution :

Variance $= s^2$

$$= \sum_{x=0}^{\infty} x^2P(X=x)-m^2$$

$$= \sum_{x=0}^{\infty} x^2\frac{e^{-m}\ m^x}{x!}-m^2$$

$$= \sum_{x=0}^{\infty}\left(x(x-1)\frac{e^{-m}\ m^x}{x!}+\frac{xe^{-m}\ m^x}{x!}\right)-m^2$$

$$= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-m}\ m^x}{x(x-1)(x-2)!}+\sum_{x=0}^{\infty}\frac{xe^{-m}\ m^x}{x!}-m^2$$

$$= \sum_{x=0}^{\infty} \frac{e^{-m}\, m^2 m^{x-2}}{(x-2)!} + m - m^2$$

$$= e^{-m} m^2 \sum_{x=2}^{\infty} \frac{m^{x-2}}{(m-2)!} + m - m^2$$

$$= e^{-m} m^2 \left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots\ldots\infty \right] + m - m^2$$

$$= e^{-m} m^2 e^m + m - m^2$$

$$= m^2 + m - m^2 = m$$

Note : Mean of a poisson distribution

= Variance

= m.

Recurrence relation for a poisson distribution

We have $P(X = x) = \frac{e^{-m} m^x}{x!},\ x = 0, 1, 2, \ldots.\infty$

$$\therefore P(X = x+1) = \frac{e^{-m} m^{x+1}}{(x+1)!}$$

$$\therefore \frac{P(X = x+1)}{P(X = x)} = \frac{e^{-m} m^{x+1}}{(x+1)!} \frac{x!}{e^{-m} m^x}$$

$$= \frac{m}{x+1}$$

$$\therefore P(X = x+1) = \frac{m}{x+1} P(X = x)$$

Example 6 : A random variable X follows poisson distribution such that P(X = 0) = P(X = 1) = a. Find the mean of the distribution. Also find the value of a.

Solution : We have $P(X = x) = \frac{e^{-m} m^x}{x!}$

Now, P(X = 0) = P (X = 1), (given)

$$\therefore \frac{e^{-m} m^0}{0!} = \frac{e^{-m} m^1}{1!}$$

$$\therefore m = 1$$

i.e., mean = 1

Again P (X = 0) = a

or, $e^{-m} = a$

or, $e^{-1} = a$

i.e. $a = \frac{1}{e}$

Example 7 : A random variable X follows poisson distribution such that P(X = 1) = 2P(X = 2). Find P(X = 0), $P(X \geq 0)$.

Solution : P(X = 1) = 2P(X = 2), given

$$\therefore \frac{e^{-m}m^1}{1!} = 2\,\frac{e^{-m}m^2}{2!}$$

or, m = 1

$$\therefore P(X=0) = \frac{e^{-1}1^0}{0!} = e^{-1} = \frac{1}{e}$$

$$P(X \geq 0)$$

$$= P(X=0) + P(X=1) + P(X=2) + \ldots.$$

$$= e^{-1} + \frac{e^{-1}}{1!} + \frac{e^{-1}}{2!} + \frac{e^{-1}}{3!} + \frac{e^{-1}}{4!} + \ldots..$$

$$= e^{-1}\left(1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{3!} + \frac{1}{4!} + \ldots..\right)$$

$$= e^{-1}\, e$$

$$= e^0 = 1$$

Example 8 : 6 coins are tossed 6400 times. Use poisson distribution to find the approximate probability of getting 6 heads r times.

Solution : Here n = 6400

p = the probability of getting 6 heads in one throw of 6 coins

$$\therefore\ m = np = \frac{6400}{2^6} = \frac{6400}{64} = 100$$

$$\therefore P(X=r) = \frac{e^{-m}m^r}{r!}$$

$$= \frac{e^{-100}100^r}{r!}$$



CHECK YOUR PROGRESS 2

Q 1: A random variable X follows Poisson distribution such that P(X = 1) = P(X = 2). Find P(X = 3)

Q 2: A random variable X follows poisson distribution such that P(X = 2) = 3 P(X = 3). Find the probability that X is at least one.

Q 3: Examine if the following statement is right.

The mean of a poisson distribution is 3, and its standard deviation is 4.

Q 4: A random variable X follows poisson distribution such that P(X = 2) = 9P(X = 4). Find the mean of X.

Q 5: A random variable X follows poisson distribution such that P(X = 2) = P(X = 3). Find the variance of X.



13.6 LET US SUM UP

- A Bernoulli trial is an experiment satisfying the following conditions.
  1. There are only two possible outcomes, viz,. success or failure (occurrence of an event is ‘success’, non-occurrence is ‘failure’).
  2. The probability of a success (or failure) remains same in each trial.
  3. Trials must be independent.
  4. Number of trials must be finite.
  5. The sum of probabilities of success and failure is 1.
  6. The probability of success vis-a-vis failure is not very low.
- Let there be n Bernoulli trials. Let the random variable X denote the number of successes in these trials. Let the probability of getting a success in a single trial be p, and that of getting a failure be q so that p + q = 1. Then $P(X = r) = {}^{n}C_{r}\, p^{r}\, q^{n-r}$

  This is known as Binomial distribution.

- The Bernoulli distribution is a special case of the binomial distribution where n = 1.
- Mean of a binomial distribution is np, and the variance is npq.
- Poisson distribution is a limiting case of the binomial distribution under the following there conditions.
  1. The number of trials n is indefinitely large, i.e., $n \to \infty$.
  2. The constant probability p of success for each trial is indefinitely small, i.e., $p \to 0$.
  3. np = m, (say), is finite.
- A random variable X is said to follow a poisson distribution if it takes non-negative values with the probability given by

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

- Mean of a poisson distribution
  = variance
  = m



13.7 FURTHER READINGS

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009
2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011
3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005
4) Probability, Random Process, Queueing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042

5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33



13.8 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: If a coin is tossed, S = {H, T}

$\therefore$ P (getting a head) $\frac{1}{2}$

P (not getting a head) $= 1 - \frac{1}{2} = \frac{1}{2}$

Let X be the random variable denoting the number of tails.

In 4 trials there may be 0 or 1 or 2 or 3 or 4 tails.

$$P(X=0) = {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0} = \frac{1}{16}$$

$$P(X=1) = {}^4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{4-1} = \frac{1}{4}$$

$$P(X=2) = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} = \frac{3}{8}$$

$$P(X=3) = {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{4-3} = \frac{1}{4}$$

$$P(X=4) = {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4} = \frac{1}{16}$$

$\therefore$ the probability distribution is

X	0	1	2	3	4
P(X)	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Ans to Q No 2: [One of the meanings of the word doublet is as follows.

'Either of two dice, each of which, when thrown, has the same number of spots on the face lying uppermost.']

If a pair of dice is thrown, there will be 36 outcomes.

S = { (1, 2), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), ..., (6, 1), (6, 2), (6, 3),

(6, 4), (6, 5), (6, 6)}

Thus n(s) = 36

Possible doublets are (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6).

i.e., the number of doublets is 6.

$\therefore$ P(getting a doublet) $= \frac{6}{36} = \frac{1}{6}$

$\therefore$ P(not getting a doublet) $= 1 - \frac{1}{6} = \frac{5}{6}$

In 3 trials, there may be 0 or 1 or 2 or 3 doublets.

$P(X = 0) = {}^3C_0 \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3 = \frac{125}{216}$

$P(X = 1) = {}^3C_1 \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^2 = \frac{25}{72}$

$P(X = 2) = {}^3C_2 \left(\frac{1}{6}\right)^2 \frac{5}{6} = \frac{5}{72}$

$P(X = 3) = {}^3C_3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 = \frac{1}{216}$

$\therefore$ the probability distribution is

X	0	1	2	3
P(X)	$\frac{125}{216}$	$\frac{25}{72}$	$\frac{5}{72}$	$\frac{1}{216}$

Ans to Q No 3: If a coin is flipped, S = {H, T}

$\therefore$ P(getting a head) $= \frac{1}{2}$

P (not getting a head) $= 1 - \frac{1}{2} = \frac{1}{2}$

Let X be the random variable denoting the number of heads.

$\therefore P(X = r) = {}^nC_r p^r q^{n-r}$

$= {}^8C_r \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{8-r} = {}^8C_r \left(\frac{1}{2}\right)^8$

$\therefore$ P (getting at least 2 heads)

$= P(X \geq 2)$

$$=1-[P(X=0)+P(X=1)]$$

$$=1-\left[{}^8C_0\left(\frac{1}{2}\right)^8+{}^8C_1\left(\frac{1}{2}\right)^8\right]$$

$$=1-\left[\frac{1}{256}+\frac{8}{256}\right]$$

$$=1-\frac{9}{256}$$

$$=\frac{247}{256}$$

Ans to Q No 4: Let X be the binomial variate.

We have mean = np = 3 (1)

variance = npq = 2 (2)

From (1) and (2), 3q = 2

$$\therefore\ q=\frac{2}{3}$$

$$\therefore\ p=1-q=1-\frac{2}{3}=\frac{1}{3}$$

Again, from (1) $\frac{n}{3}=3$

$$\therefore\ n=9$$

$\therefore$ the distribution is

$$P(X=r)={}^9C_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{9-r},\qquad r = 0, 1, 2, 3, ... , 9.$$

Ans to Q No 5: We have

9p (x = 4) = p(x = 2)

or, 9 ${}^6C_4\ p^4\ q^{6-4}$ = ${}^6C_2\ p^2\ q^{6-2}$, (n = 6 given)

or, $9p^2 = q^2$, (${}^6C_4 = {}^6C_2$)

or, 3p = q

or, 3p = 1 – p

or, 4p = 1

$$\therefore\ p=\frac{1}{4}\qquad\therefore\ q=1-\frac{1}{4}=\frac{3}{4}$$

$$\therefore\ \text{mean}=np=6\times\frac{1}{4}=\frac{3}{2}$$

variance $= npq = 6 \times \frac{1}{4} \times \frac{3}{2} = \frac{9}{4}$

Ans to Q No 6: We have

mean = 2 (standard deviation), (given)

or, $np = 2\sqrt{npq}$

or, $n^2 p^2 = 4npq$

or, np = 4q (1)

Again, $P(X = 0) = \frac{1}{16}$ (given)

$\therefore q^n = \frac{1}{16}$

or $q^n = \left(\frac{1}{2}\right)^4$ or $q^n = \left(\frac{1}{4}\right)^2$

$\therefore$ either $q = \frac{1}{2},\ n = 4$ or $q = \frac{1}{4},\ n = 2$

If $q = \frac{1}{2},\ n = 4$ then from (1)

$4p = 4 \times \frac{1}{2}$ or $p = \frac{1}{2}$. $\therefore q = 1 - p = \frac{1}{2}$

If $q = \frac{1}{4},\ n = 2$ then from (1)

$2p = 4 \times \frac{1}{4}$ or 2p = 1 or $p = \frac{1}{2}$

$\therefore\ q = 1 - \frac{1}{2} = \frac{1}{2}$ which is not true.

$\therefore\ q = \frac{1}{4},\ n = 2$ is not admissible

$\therefore n = 4,\ q = \frac{1}{4},\ p = \frac{1}{2}$

P (X is at most 1)

= P(X = 0) + P(X = 1)

$= {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0} + {}^4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{4-1}$

$= \left(\frac{1}{2}\right)^4 + 4 \times \left(\frac{1}{2}\right)^4$

$$= \frac{1}{16} + \frac{4}{16}$$

$$= \frac{5}{16}$$

Ans to Q No 7: If a coin is flipped, S = {H, T}

$\therefore$ P(getting a head) $= \frac{1}{2}$

P (not getting a head) $= 1 - \frac{1}{2} = \frac{1}{2}$

$\therefore \; p = \frac{1}{2}, \; q = \frac{1}{2}$

Let X be the random variable denoting the number of heads.

In 8 trials there may be 0 or 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 heads.

$\therefore$ P (getting at least 7 heads)

$$= P(X = 7) + P(X = 8)$$

$$= {}^8C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^{8-7} + {}^8C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^{8-8}$$

$$= 8 \times \frac{1}{2^8} + \frac{1}{2^8}$$

$$= \frac{8}{256} + \frac{1}{256}$$

$$= \frac{9}{256}$$

Ans to Q No 8: We have np = 6 (1)

$$\sqrt{npq} = \sqrt{2}$$

$\therefore \; npq = 2$

or 6q = 2

$\therefore \; q = \frac{1}{3}$

$\therefore \; p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3}$

$\therefore$ from (1), $\frac{2n}{3} = 6$ or n = 9

$\therefore P(X = r) = {}^9C_r \left(\frac{2}{3}\right)^r \left(\frac{1}{q}\right)^{q-r}$, r = 0, 1, 2, 3, 4, 5, 6,7, 8, 9

Ans to Q No 9: We have np = 16, (1)

npq = 8 (2)

$\therefore 16q = 8$ or $q = \frac{1}{2}$

$\therefore\ p = 1 - q = 1 - \frac{1}{2} = \frac{1}{2}$

$\therefore$ from (1), $\frac{n}{2} = 16$ or n = 32

(i) $P(X = 0) = {}^{32}C_0 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^{32-0}$

$= \frac{1}{2^{32}}$

(ii) $P(X = 1) = {}^{32}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{32-1}$

$= 32 \cdot \frac{1}{2^{32}} = \frac{2^5}{2^{32}} = \frac{1}{2^{27}}$

(iii) $P(X \geq 2)$

$= 1 - [P(X = 0) + P(X = 1)]$

$= 1 - \left(\frac{1}{2^{32}} + \frac{1}{2^{27}}\right)$

Ans to Q No 10: (i) Mean = 6, SD = 3

$\therefore$ variance = 9

We know that for a binomial distribution mean > variance

In this case mean = 6, variance = 9

$\therefore$ the statement is not true.

(ii) Mean = 5

Variance = 2

We know that for a binomial distribution, mean > variance.

In this case mean = 5, variance = 2.

$\therefore$ the statement is true.

Check Your Progress 2

Ans to Q No 1: We know that, for a poisson distribution,

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

Now P(X=1) = P(X=2), (given)

$$\frac{e^{-m}m^1}{1!}=\frac{e^{-m}m^2}{2!}$$

$\therefore$ m = 2

$$\therefore\ P(X=3)=\frac{e^{-2}\,2^3}{3!}=\frac{4e^{-2}}{3}=\frac{4}{3e^2}$$

Ans to Q No 2: We have

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

P (x = 2) = 3 P (x = 3)

or, $\frac{e^{-m}\,m^2}{2!}=3\,.\,\frac{e^{-m}\,m^3}{3!}$

or, m = 1

$$\therefore\ P(X\geq 1)$$

$$=1-P(X=0)$$

$$=1-\frac{e^{-1}1^0}{0!}$$

$$=1-e^{-1}\ =1-\frac{1}{e}\ =\frac{e-1}{e}=0.632$$

Ans to Q No 3: We know that the mean of a poisson distribution = variance.

$\therefore$ the given statement is not true.

Ans to Q No 4: We have $P(X=x)=\frac{e^{-m}\,m^x}{x\,!}$

Now P(x = 2) = 9 P(x = 4)

or, $\frac{e^{-m}\,m^2}{2\,!}=9\,.\,\frac{e^{-m}\,m^4}{4\,!}$

or, $\frac{m^2}{2}=\frac{3m^4}{8}$

or, $\frac{3m^2}{4}=1$

or, $m^2=\frac{4}{3}$

$\therefore\ m = \frac{2}{\sqrt{3}}$

$\therefore$ mean = $m = \frac{2}{\sqrt{3}}$

Ans to Q No 5: We have

P(x = 2) = P(x = 3)

or, $\frac{e^{-m} m^2}{2!} = \frac{e^{-m} m^3}{3!}$

or, m = 3

$\therefore$ variance = m = 3



13.9 MODEL QUESTIONS

Q 1: Name the discrete probability distribution for which

(i) mean is greater than the variance,

(ii) mean is equal to the variance.

Q 2: Derive poisson distribution as a limiting form of a binomial distribution.

Q 3: 8 coins are thrown simultaneously. Find the probability of obtaining at least 6 heads.

Q 4: Let X be a binomial variate such that mean = 1, variance = 0.75. Find (i) P(x = 0), (ii) P (x > 0).

Q 5: Write down the conditions under which a binomial distribution tends to the poisson distribution.

Q 6: Let X be a poisson variate such that $P(X = 2) = \frac{2}{3} P(X = 1)$.

Find P(x = 0)

Q 7: If the sum of the mean and variance of a binomial distribution for 5 trials is 1.8, find the distribution.

Q 8: The mean and variance of a binomial distribution are 4 and $\frac{4}{3}$ respectively. FInd $P(X \geq 1)$.

*** ***** ***

UNIT 14 : THEORETICAL DISTRIBUTION II

UNIT STRUCTURE

14.1 Learning Objectives

14.2 Introduction

14.3 ContinuousProbability Distribution

14.4 Normal Distribution

14.4.1 Properties of Normal Distribution

14.4.2 Standard Normal Variate

14.5 Let Us Sum Up

14.6 Further Readings

14.7 Answers To Check Your Progress

14.8 Model Questions

14.1 LEARNING OBJECTIVES

After going through this unit, you will able to learn

- continuous probability distribution and probability density function
- normal distribution
- standard normal variate
- properties of normal distribution.

14.2 INTRODUCTION

Normal distribution is a continuous distribution. It is a limiting case of the binomial distribution. It has immense application. It is the most popular and commonly used distribution. It was discovered by Abraham de Moivre (1667–1754) in 1733.

14.3 CONTINUOUS PROBABILITY DISTRIBUTION

If a random variable is a continuous variable, its probability distribution is called a continuous probability distribution.

A continuous probability distribution cannot be expresed in tabular form. An equation or formula is used to describe a continuous probability

distribution. The equation used to describe a continuous probability distribution is called a probability. density function (p.d.f.)

14.4 NORMAL DISTRIBUTION

Normal distribution is derived from the binomial distribution under two conditions, viz.

(i) n, the number of trials, is sufficiently large, i.e. $n \to \infty$

and (ii) neither p nor q is very small.

A random variable X is said to follow normal distribution, if and only if, its probability density function is given by

$$f(x) = \frac{1}{s\sqrt{2p}}\, e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

Where x is the real value of X, i.e. $-\infty < X < \infty$

The variable X is said to be distributed normally with mean m and variance s^2 symbolically we write $X \sim N(m, s^2)$. It is read as 'X is distributed as $N(m, s^2)$'.

Note :

1. The probability density function given by (1) has two parameters, viz., m and s .
   m is any real number, i.e. $-\infty < m < \infty$ whereas s is a positive real number, i.e., $s > 0$.
2. We know that probability can never be negative.
   $\therefore f(x) \geq 0, \forall x$
3. If $m = 0,\ s = 1$ the p.d.f is given by

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

In this case $X \sim N(0, 1)$. In this case X is called the standard normal variate, and the p.d.f given by (2) is called standard normal distribution.

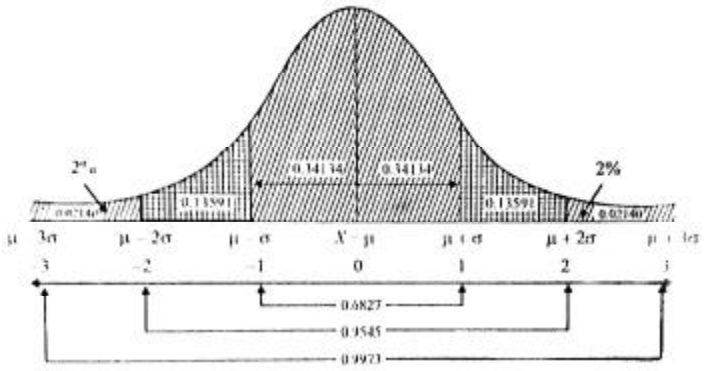
14.4.1 Properties of Normal Distribution

Normal distribution has many properties, some of which are given below.

1. The equation of the normal probability curve is

$$Y = \frac{1}{s\sqrt{2p}}\, e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

It is of the shape given below. It is bell-shaped.



[Source : **Biostatistics** by Dr. P N Arora and Dr. P K Malhan (Himalya Publishing House, 2010)]

2. The curve is symmetrical about the line $x = \infty$, and x lies between $-\infty$ and $+\infty$

3. At $x = m$, the ordinate is maximum, and it is $Y = \frac{1}{s\sqrt{2p}}$

4. The total area under the normal curve within the limits $-\infty$ to ∞ is 1, i.e.,

$$\frac{1}{s\sqrt{2p}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}\, dx = 1$$

5. The percentage distribution of area under the normal curve given below, and is shown in the figure given above.
   (i) About 68% of the area falls between $m - s$ and $m + s$
   (ii) About 95.5% of the area falls between $m - 2a$ and $m + 2s$.
   (iii) About 99.7% of the area falls between $m - 3s$ and $m + 3s$.
6. The curve has a single peak, i.e., it is unimodal.
7. The distribution is symmetrical. Mean = median = mode = m .
8. The curve is asymptotic, i.e., it meets the x-axis at both the

ends at infinity. Theoretically the curve never touches the x-axis.

9. The points of inflexion are $x = m - s$ and $x = m + s$. [Inflexion points are where the function changes concavity.]
10. Coefficient of skewness = 0, and kurtosis = 3.
11. Mean deviation about mean $\simeq \frac{4}{5}s$
12. $Q.D. \simeq \frac{2}{3}s$

14.4.2 Standard Normal Variate

Let a random variable X be distributed as $N(m, s^2)$. The location and shape of the normal curve depends on m, and s, Where $-\infty < m < \infty$, and $s > 0$. Therefore, no master table for the area can be prepared. This hindrance can be removed if we introduce a new variable Z defined by.

$$Z = \frac{X - m}{s}$$

In this case Z is called the standard normal variate.

Now we prove that the mean of Z is 0, and the standard deviation is 1. Before that we state some related properties.

We know that mean of a random variable X is given by

mean $= E(X) = \Sigma x_i p_i$

We prove that

(i) E(a) = a; a is a constant

(ii) E (aX + b) = aE(x) + b; a, b are constants

(iii) $E\left(X - \overline{X}\right) = 0$

Proof : (i) $E(a) = \Sigma a p_i$

$= a\ \Sigma p_i$

= a, **Q** $\Sigma p_i = 1$

(ii) $\Sigma(aX + b) = \Sigma(ax_i + b)\, p_i$

$= a\,\Sigma x_i p_i + b\,\Sigma p_i$

$= a\,E(X) + b,$ **Q** $\Sigma p_i = 1$

(iii) $E\left(X-\bar{X}\right)$

$=E(X)-E(\bar{X})$

$=\bar{X}-\bar{X}$, **Q** $\bar{X}$ is constant

= 0

Also, we note that

(iv) Var (aX) = a^2 Var (X), where a is a constant

(v) Var ($aX \pm b$) = a^2 Var (X), where a, b are constants.

Now we prove that the mean of Z is 0, and the standard deviation is 1.

Mean = E (Z)

$=E\left(\frac{X-m}{s}\right)$

$=\frac{1}{s}E(X-m)$, property (ii)

$=\frac{1}{s}\left[E(X)-m\right]$, property (ii)

$=\frac{1}{s}[m-m]$

= 0

Variance = Var (Z)

$=Var\left(\frac{X-m}{s}\right)$

$=\frac{1}{s^2}\text{var}(X-m)$, property (v)

$=\frac{1}{s^2}\text{var } X$, property (iv)

$=\frac{1}{s^2}.s^2$

= 1

Thus, $Z=\frac{X-m}{s}$ is distributed as N (0, 1), i.e., Z ~ N(0, 1).

All the properties for a normal variable hold good for standard normal variable $(m=0,\ s=1)$

Let Z be a standard normal variate N(0, 1). The area under the standard normal curve is given by a table which is addended at the end of this chapter.

We know that the total area under the normal curve within the limits $-\infty$ to ∞ is 1. The standard normal curve is symmetrical about Z = 0. Therefore, the area to the left of z = 0 is 0.5, and the area to the right of z = 0 is 0.5.

Suppose, we are to find the area between z = 0, and z = 1.89, i.e, we are to find out $P(0 \le Z \le 1.89)$.

First we find 1.8 at the left of the table and single out that row.

Next we find 0.09 at the top of the table and single out that column.

Then we obtain the entry where the row for 1.8 and the column for 0.09 meet. the entry is 0.47062.

Thus $P(0 \le Z \le 1.89)$

= (the are a from z = 0 to z = 1.89)

= 0.47062.

Example 1 : Find the probability of the standard normal variate z ehen z is greater than 1.09.

Solution : We know that the total area under the normal curve within the limits $-\infty$ to ∞ is 1. Therefore, the area to the right of z = 0 is 0.5, as the curve is symmetrical about z = 0.

$\therefore$ P(Z > 1.09) = 0.5 – 0.36214 = 0.13786

Example 2 : Find the probability of the standard normal variate z when z is less than 1.09.

Solution : We know that the total area under the normal curve within the limits $-\infty$ to $+\infty$ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the right or z = 0 is 0.5.

Again, from the table, we get that the area from z = 0 to z = 1.09 is 0.36214.

$\therefore$ P(z < 1.09)

= (area to the left of z = 0) + (area from z = 0 to z = 1.09)

= 0.5 + 0.36214

= 0.86214

Example 3 : Find P(0.5 < z < 1.09)

Solution : We know that the area to the right of z = 0 is 0.5. From the table, the area from z = 0 to z = 0.5 is 0.19146. Again, the area from z = 0 to z = 1.09 is 0.36214.

∴ P(0.5 < z < 1.09)

= (Area from z = 0 to z = 1.09) – (area from z = 0 to x = 0.5)

= 0.36214 – 0.19146

= 0.17068

Example 4 : A large number of measurements is normally distributed with mean 65.5 cm and standard deviation 6.2 cm. Find the percentage of measurements that fall between 54.8 cm and 68.8 cm.

Solution : Let X be the normal variate showing the large number of measurements.

We have $m = 65.5,\ s = 6.2$

Now, $Z = \frac{X - m}{s} = \frac{X - 65.5}{6.2}$

When, X = 54.8 $Z = \frac{54.8 - 65.5}{6.2} = \frac{-10.7}{6.2} = -1.7$

When, X = 68.8, $Z = \frac{68.8 - 65.5}{6.2} = 0.5$

Thus, when X lies between 54.8 and 68.8, Z lies between –1.7 and 0.5

Now P(–1.7 < X < 0.5)

= P(–1.7 < Z < 0) + P(0 < Z < 0.5)

= 0.45543 + 0.19146

= 0.64689 = 64.69%

∴ the percentage of measurements is 64.69%

Example 5 : For the distribution of marks of candidates in an examination, mean is 14 and standard deviation is 2.5. Assuming the normality of the distribution, find the probability that a randomly

selected student will score above 15.

Solution : Let X be the normal variate showing the scores of candidates.

We have $m = 14,\ s = 2.5$

Now, $Z = \dfrac{X - m}{s} = \dfrac{X - 14}{2.5}$

When X=15 $Z = \dfrac{15-14}{2.5} = \dfrac{1}{2.5} = 0.4$

$\therefore$ P(x > 15) = P(z > 0.4)

= 0.5 – (Area between z = 0 and Z = 0.4)

= 0.5 – 0.15542

= 0.34458.



CHECK YOUR PROGRESS 1

Q 1: Find (i) P (Z > 2.58), (ii) P(Z < 1.96), (iii) P (Z > –3)

Q 2: For a certain normal distribution, mean is 50, and standard deviation is 5. Find $P(45 \le X \le 55)$, X being a normal variate.

Q 3: A normal variate X has mean 100 and standard deviation 4. Find the probability that X lies between 92 and 10.8

Q 4: The distribution of scores of students in an examination is normal. If the mean is 45, and the standard deviation is 5, find the percentage of students who scored between 40 and 50.

Q 5: The marks obtained in a certain examination follow normal distribution with mean 45, and standard deviation 10. If 1000 students appear at the examination, calculate the number of students scoring (i) less than 30 marks, and (ii) more than 80 marks.



14.5 LET US SUM UP

- Normal distribution is a continous distribution. It is a limiting case of the binomial distribution.
- Normal distribution is derived from the binomial distribution under two conditions, viz.,
  - (i) n, the number of trials, is sufficiently large, i.e., $n \to \infty$ and
  - (ii) neither p nor q is very small.

  A random variable X is said to follow normal distribution, if and only if, its probability density function is given by

  $$f(x) = \frac{1}{a\sqrt{2p}}\, e^{-\frac{1}{2}\left(\frac{x-m}{a}\right)^2}$$

  Where x is the real value of X, i.e., $-\infty < X < \infty$

  The variable X is said to be distributed normally with mean m and variance s^2. Symbolically, we write $X \sim N(m,\ s^2)$. It is read as 'X is distributed as $N(m,\ s^2)$.
- The normal curve is bell-shaped, and the curve is symmetrical about the line $x = m$, and x lies between $-\infty$, and $+\infty$.
- The total area under the normal curve within the limits $-\infty$ to ∞ is 1.
- If a new variable Z is defined by $Z = \frac{X - m}{s}$, Z is called the standard normal variate. The mean of Z is 0, and the standard deviation is 1.



14.6 FURTHER READING

1) Statistical Methods by N.G. Das, McGraw Hill Education (India) Private Ltd. 2009
2) Probability and Statistics by G.Shanker Rao, University Press (India) private Ltd. 2011
3) Introduction to probability and Statistics by Seymour Lipschutz, John J.Schiller, Tata McGraw Hill Education private Ltd. 2005

4) Probability, Random Process, Queueing theory and Statistics by Dr. S.Rajasekaran, Prof.A.Chandrasekaran & G.Kavitha, Dhanam Publications, Chennai-60042

5) Probability and Statistics by G.Balaji, G.Balaji Publishers, Chennai-33



14.7 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1

Ans to Q No 1: (i) We know that tha total area under the normal curve within the limits $-\infty$ to $+\infty$ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the night of z = 0 is 0.5. Again, from the table, we get that the area from z = 0 to z = 2.58 is 0.49506

$\therefore$ P (z > 2.58)

= (area to the right of z = 0) – (area from z = 0 to z = 2.58)

= 0.5 – 0.49506

= 0.00494

(ii) We know that the total area under the normal curve within the limits $-\infty$ to $+\infty$ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the right of z = 0 is 0.5. Again from the table, we get that the area from z = 0 to z = 1.96 is 0.47500.

$\therefore$ P (z < 1.96)

= (area to the left of z = 0) + (area from z = 0 to z = 1.96)

= 0.5 + 0.47500

= 0.97500

(iii) We know that the total area under the normal curve within the limits $-\infty$ to ∞ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the right of z = 0 is 0.5.

Now, P(Z > –3)

= area from z = –3 to z = 0) + (area to the right of z = 0)

= (area from z = 0 to z = 3) + (area to the right of z = 0)

= 0.49865 + 0.5

= 0.99865

Ans to Q No 2: X is a normal variate.

We have $m = 50,\ s = 5$

Now, $Z = \frac{X - m}{s} = \frac{X - 50}{5}$

When X = 45 $Z = \frac{45 - 50}{5} = -1$

When X = 55, $Z = \frac{55 - 50}{5} = 1$

Thus, when X lies between 45 and 55, Z lies between –1 and 1.

Now $P(-1 \le Z \le 1)$

$= P(-1 \le Z \le 0) + P(0 \le Z \le 1)$

= 0.34134 + 0.34134

= 0.68268

Ans to Q No 3: Do yourself.

Ans to Q No 4: Do yourself.

Ans to Q No 5: Let X be the normal variate showing marks.

We have $m = 45,\ s = 10$

Now $Z = \frac{X - m}{s} = \frac{X - 45}{10}$

When X = 30, $Z = \frac{30 - 45}{10} = \frac{-15}{10} = -1.5$

When X = 80, $Z = \frac{80 - 45}{10} = \frac{35}{10} = 3.5$

(i) Now $P(X \le 30)$

$= P(z \le .1.5)$

= (area to the left of z = 0) – (area from z = – 1.5 to z = 0)

(area to the left of z = 0) – (area from z = 0) to z = 1.5)

= 0.5 – 0.43319

= 0.06681

∴ the required number of students

= 1000 x 0.06681

= 67 (approximately)

(ii) $P(Z \geq 80)$

$= P(Z \geq 3.5)$

= (area to the right of z = 0) – (area from z = 0 to z = 3.5)

= 0.5 – 0.49977

= 0.00023

Now 1000 x 0.00023 = 0.23

∴ there are no students scoring more than 80 marks.



14.8 MODEL QUESTIONS

Q 1: Define standard normal variate and write down the probability density function of this variate.

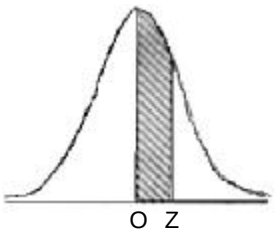
Q 2: State the conditions under which the normal distribution is obtained as a limiting case of the binomial distribution.

Q 3: Find (i) $P(Z \leq -1.96)$ (ii) $P(Z \leq 1.96$ (iii) $P(Z > 2.58)$

Q 4: A normal variate X has mean 50 and standard deviation 5. Find $P(40 \leq X \leq 60)$.

Q 5: The mean of a normal distribution is 50 and the standard deviation is 4. What percentage of total frequency lies between 42.16 and 57.84?

Q 6: Let a continuous variable X follow a normal distribution with mean 12 and standard deviation 2. What is the probability that the value of X selected at random lies between 11 and 14 ?



Area under the Standard Normal Curve

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04383	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910-	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31372
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34849	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	:37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251'	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41198	.41309	.41466	.41621	.41774
1.4	.41924	.42073-	.42220	.42364	.42507	.42647	.42785	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47784	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899

2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.5	.49379	.49396	.49413	.49430	.49446	.49461'	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632.	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49861
3.0	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49893	.49896	.49900
3.1	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929
3.2	.49931	.49934	.49936	.49938	.49940	.49942	.49944	.49946	.49948	.49950
3.3	.49952	.49953	.49955	.49957	.49958	.49960	.49961	.49962	.49964	.49965
3.4	.49966	.49968	.49969	.49970	.49971	.49972	.49973	.49974	.49975	.49976
3.5	.49977	.49978	.49978	.49979	.49980	.49981	.49981	.49982	.49983	.49983
3.6	.49984	.49985	.49985	.49986	.49986	.49987	.49987	.49988	.49988	.49989
3.7	.49989	.49990	.49990	.49990	.49991	.49991	.49992	.49992	.49992	.49992
3.8	.49993	.49993	.49993	.49994	.49994	.49994	.49994	.49995	.49995	.49995
3.9	.49995	.49995	.49996	.49996	.49996	.49996	.49996	.49996	.49997	.49997
4.0	.49997	-	-	-	-	-	-	-	-	-
4.5	.499997	-	-	-	-	-			-	-
5.0	.4999997	-	-	-	-	-	-	-	-	-

[Source : Basic statisties by B. L. Agarwal (New Age International Publishers, 2009)]

*** ***** ***



REFERENCES

1. Statics : B. C. Das and B. N. Mukerjee U. N. Dhur and Pus.
2. A Text Book an statics by R. S. Verma.
3. An introduction to commercial Mathematics and statistics - Prof. Ajanta Mazumder, Dr. (Mrs) Runjan Phookun. S. Chand and Company Lt. R New Delhi.
4. Business Statistics - Dr. Nityananda Barman, Assam Book Depot.
5. Elementary Statistics by H.C. Saxena, S.Chand & company Ltd., 2002
6. Basic Statistics by B.L. Agarwal, New Age International Publishers, 2009
7. Biostatistics by Dr. P N Arora and Dr. P K Malhan, Himalaya publishing House, 2010
8. Fundamental of Statistics Volume one by A.M. Gun, M.K. Gupta, B. Dasgupta, world press 2005
9. Fundamental of Mathematical Statistics by S.C.Gupta, V.K.Kapoor by Sultan Chand & Sons, 2009

Name of the Paper: **Course Code:**

Centre for Internal Quality Assurance (CIQA)
Krishna Kanta Handiqui State Open University
City Office, Housefed Complex, Guwahati- 06

Learner's Feedback on Course

Dear Learner,

Regarding the course as mentioned above, we would like to know your opinions and comments so as to improve the quality of self learning materials in future. Please respond to the following statements by ticking the number you feel most reflect your opinion. After completion of the additional comments, please detach the page and send/mail the same to us at the address given below.

The Director, Centre for Internal Quality Assurance, KKHandiqui State Open University
Housefed Complex, Dispur, Guwahati- 781006
(E-mail id: ciqa@kkhsou.in)

1. Approximately how many hours did you spend for studying the units in the course? | | | |
2. Please give your opinions (by √ mark) to the following items based on your reading of the block:

Sl. No.	Statements	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
I.	The SLMs of the course fulfil the learning objectives set out in the beginning of each unit					
II.	The units of the SLM could satisfy my academic needs and expectations					
III.	The Unit writers have excellent knowledge about the course contents					
IV.	Language and contents of the units were presented at a level which I could readily understand					
V.	Ample opportunity for participation in the activities provided in the units					
VI.	Used enough Illustrations (Diagrams, tables etc.) for conceptual clarity					
VII.	Quality of content is engaging, relevant, and up-to-date					
VIII.	The self check questions are very helpful					
IX.	The Possible/model questions and the answers to check my progress have benefited me a lot					

Additional Comments: *(Please feel free to provide your open comments)*

1. Which aspects of the SLM, according to you, worked well?

   ---

2. What sort of changes/improvements do you feel KKHSOU could implement to improve the overall quality of the SLM?

Thank you for taking the time to complete this form.

******