MODULE-1

PROBABILITY DISTRIBUTIONS

Introduction

The three main components of statistics are collection of data, analysis of data and inference. The scope of statistics is very vast and finds its applications in almost all spheres. Statistical methods guides us in any scientific enquiry and leans heavily on probability theory.

Basic Definitions and concepts

Experiment is a process by which measurement or observation is obtained.

Random experiment : An experiment which does not result in the same out come when performed under the same conditions.

Eg : 1. Tossing of a fair coin.

2. Throwing of a fair die.

Sample space : The set of all possible outcomes of a random experiment is called a sample space usually denoted by S.
Eg 1. S = {H,T} is the sample space in tossing of a single coin.
2. S = {HH,TH,HT,TT} is the sample space obtained in tossing of two coins.
3. S = {1,2,3,4,5,6} in throwing of a die.

Event : Any event is a subset of an appropriate sample space.

Exhaustive event : An event consisting of all the various possibilities is called an exhaustive event.

Mutually exclusive events : Events are said to be mutually exclusive if and only if one of them can take place at a time, in other words happening of one event prevent the simultaneous happening of the others.
Eg: In the tossing of a coin, either heads or tails may turn up but not both.

Independent events: Two or more events are said to be independent if the happening or non happening of one event does not prevent the happening or non happening of the others.

Eg: 1.When two coins are tossed the event of getting head is an independent event as both the coins can turn out heads.

Probability is the measure of the chance with which we can expect the event to occur. If an event A can occur in **m** different ways out of number of **n** possible ways, all of which are equally likely then the probability of the event A denoted by P(A) is given by $P(A) = \frac{m}{n}$.

Note: 1) Since $0 \leq m \leq n \Rightarrow$ P (A) is non negative and $0 \leq P(A) \leq 1$

2) If $P(A) = 0$ then A is called a null event.

3) If $P(A) = 1$ then A is called a sure event.

4) $P(A) + P(\overline{A}) = 1$, the set of all unfavourable events is denoted by $\overline{A}$.

Addition theorem of probability

The probability of the happening of one or the other mutually exclusive events is equal to the sum of the probabilities of the two events.

That is, if A, B are two mutually exclusive events then,

$$\mathbf{P(A \text{ or } B) = P(A) + P(B)}$$

Product theorem of probability

If a compound event is made up of a number of independent events, the probability of the happening of the compound event is equal to the product of the probabilities of the independent events .

That is, if A, B are independent events then,

$$\mathbf{P(A \text{ and } B) = P(A).P(B)}$$

RANDOM VARIABLE & PROBABILITY DISTRIBUTION

Introduction

When we consider an experiment of tossing two coins. The sample points or out comes are qualitative and are described by their attributes though the probabilities attached to each of these outcomes are quantitative. We now associate to each sample point a real number called the **random variable.**

Random variable

Definition: Let S be a sample space then $\mathbf{X : S \rightarrow R}$ defined over the elements of S is called a **random variable.** A random variable is a function that assigns a real number to every sample point in the sample space of a random experiment. Random variables are usually denoted by X, Y, Z ... and it may be noted that different random variables may be associated with the sample space S. The set of all real numbers of a random variable X is called the range of X.

Eg:

1. While tossing a coin, suppose that the value 1 is associated for the outcome 'head' and 0 for the outcome tail. We have the sample space $S = \{H, T\}$ and if X is the random variable , then $X(H) = 1$ & $X(T) = 0$.

   Range of $X = \{0,1\}$

2. Suppose a coin is tossed twice, we shall associate two different random variables X, Y as follows, where we have the sample space

   $S = \{HH, HT, TH, TT\}$

   $X =$ Number of 'heads' in the outcome.

The association of the elements in S to X is as follows.

Out come	HH	HT	TH	TT
Random variable X	2	1	1	0

Range of $X = \{0,1,2\}$

Suppose Y = Number of 'tails' in the outcome, we have

Out come	HH	HT	TH	TT
Random variable Y	0	1	1	2

Range of $Y = \{0,1,2\}$

<u>Discrete and Continuous random variables</u>

If a random variable takes finite or countably infinite number of values then it is called a discrete random variable.

Discrete random variable will have finite or countably infinite range.

Eg: 1. Tossing of a coin and observing the outcome.

2. Throwing a 'die' and observing the numbers on the face.

If a random variable takes non countable number of values then it is called a non discrete or continuous random variable.

Eg:

1. Weight of articles
2. Length of nails produced by machine.

Generally counting problems correspond to discrete random variables and measuring problems lead to continuous random variables.

<u>Probability function and Discrete probability distribution</u>

If for each value x_i of a discrete random variable X, we assign a real number $p(x_i)$ such that

i) $p(x_i) \geq 0$ ii) $\sum_i p(x_i) = 1$ then the function $p(x)$ is called a probability function.

If the probability that X take the values x_i is p_i, then $P(X = x_i) = p_i$ or $p(x_i)$.

The set of values $[x_i\ ,\ p(x_i)]$ is called a discrete (finite) probability distribution of the discrete random variable X. The function P(X) is called the probability density function (p.d.f) or the probability mass function (p.m.f)

The distribution function f(x) defined by $f(x) = P(X \leq x) = \sum_{i=1}^{n} P(x_i)$, x being an integer, is called the cumulative distribution function (c.d.f).

The mean and variance of the discrete probability distribution are defined as follows

Mean $\mu = \sum_i x_i \,.\, p(x_i)$

Variance $V = \sum_i (x_i - \mu)^2 . p(x_i)$

$V = \sum_i x_i^2 . p(x_i) - [\sum_i x_i . p(x_i)]^2$

$V = \sum x_i^2 . p(x_i) - \mu^2$

Standard deviation $\sigma = \sqrt{V}$

Problems

1. A coin is tossed twice. A random variable X represent the number of heads turning up.

Find the discrete probability distribution for X. Also find its mean and variance.

Sol: Here $S = \{HH, HT, TH, TT\}$

$X = \{0,1,2\}$

Now, $P(HH) = \frac{1}{4}$ $P(HT) = \frac{1}{4}$ $P(TH) = \frac{1}{4}$ $P(TT) = \frac{1}{4}$

$$P(X = 0, \text{ i.e., no head turning up}) = P(TT) = \frac{1}{4}$$

$$P(X = 1, \text{ i.e., one head turning up}) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X = 2, \text{ i.e., two heads turning up}) = P(HH) = \frac{1}{4}$$

The discrete probability distribution for X is as follows

X=x_i	**0**	**1**	**2**
$p(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

We observe that, $p(x_i) > 0$ and $\sum_i p(x_i) = 1.$

Mean , $\mu = \sum_i x_i \,.\, p(x_i) = (0).\left(\frac{1}{4}\right) + (1).\left(\frac{1}{2}\right) + (2).\left(\frac{1}{4}\right) = \frac{1}{2} + \frac{1}{2}$

$\boldsymbol{\mu = 1}$

Variance, $V = \sum_i (x_i - \mu)^2 . p(x_i)$

$V = (0-1)^2 . \frac{1}{4} + (1-1)^2 . \frac{1}{2} + (2-1)^2 . \frac{1}{4}$

$V = \frac{1}{4} + 0 + \frac{1}{4}$

$\mathbf{V = \frac{1}{2}}$

2. A random experiment of tossing a 'die' twice is performed. Random variable X on this sample space is defined to be the sum of the two numbers turning up on the toss. Find the discrete probability distribution for the random variable X and compute the corresponding mean and standard deviation.

Sol: Here $S = \{(x, y)$ where $x = 1,2, \dots, 6\ ; y = 1,2, \dots, 6\}$

i.e., $S = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), \dots, (6,5), (6,6)\}$

Number of elements in S=36.

The set of values of the random variable X defined as sum of two numbers on the face of the 'die' are $\{2,3,4,5, \dots, 10,11,12\}$

Now, $p(x_1) = \frac{n(E)}{n(S)} = \frac{1}{36}\ ; p(x_2) = \frac{2}{36}\ ; \dots$

The discrete probability distribution for X is as follows

$\mathbf{X = x_i}$	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$p(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Here $p(x_i) > 0$ and $\sum_i p(x_i) = 1$.

Mean $\mu = \sum_i x_i \,.\, p(x_i)$

$$\mu = (2).\left(\frac{1}{36}\right) + (3).\left(\frac{2}{36}\right) + \dots + (11).\left(\frac{2}{36}\right) + (12).\left(\frac{1}{36}\right)$$

$$\boldsymbol{\mu = \frac{252}{36} = 7}$$

Variance, $V = \sum_i (x_i - \mu)^2 . p(x_i)$

$$V = (2-7)^2.\frac{1}{36} + (3-7)^2.\frac{2}{36} + \dots + (12-7)^2.\frac{1}{36}$$

$$\mathbf{V = \frac{210}{36} = \frac{35}{6}}$$

Std. deviation , $\sigma = \sqrt{V}$

$$\boldsymbol{\sigma = \sqrt{\frac{35}{6}} = 2.42}$$

3. Show that the following distribution represents a discrete probability distribution. Find the mean and variance

x	10	20	30	40
p(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Sol: We observe that $p(x) > 0$ and $\sum p(x) = 1$.

Mean $\mu = \sum x\,.\,p(x)$

$$\mu = (10).\tfrac{1}{8} + (20).\tfrac{3}{8} + (30).\tfrac{3}{8} + (40).\tfrac{1}{8}$$

$$\mu = \frac{10+60+90+40}{8} = \frac{200}{8}$$

$$\boldsymbol{\mu = 25}$$

Variance, $V = \sum (x-\mu)^2.\,p(x)$

$$V = (10-25)^2.\tfrac{1}{8} + (20-25)^2.\tfrac{3}{8} + (30-25)^2.\tfrac{3}{8} + (40-25)^2.\tfrac{1}{8}$$

$$\mathbf{V = \frac{600}{8} = 75}$$

4. Find the value of k such that the following distribution represents a finite probability distribution. Hence find its mean and standard deviation. Also find $p(x \le 1)$, $p(x > 1)$ and $p(-1 < x \le 2)$

x	-3	-2	-1	0	1	2	3
p(x)	k	2k	3k	4k	3k	2k	k

Sol: We must have $p(x) \ge 0$ for all x and $\sum p(x) = 1$. The first condition is satisfied if $k \ge 0$. Since $\sum p(x) = 1$

$$k + 2k + 3k + 4k + 3k + 2k + k = 1$$

$$16k = 1$$

$$\therefore\ \mathbf{k = \frac{1}{16}}$$

The discrete probability distribution is as follows

x	-3	-2	-1	0	1	2	3
p(x)	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Mean $\mu = \sum x\,.\,p(x)$

$$\mu = (-3).\frac{1}{16} + (-2).\frac{2}{16} + (-1).\frac{3}{16} + (0).\frac{4}{16} + (1).\frac{3}{16} + (2).\frac{2}{16} + (3).\frac{1}{16}$$

$$\boldsymbol{\mu = 0}$$

Variance, $V = \sum (x-\mu)^2.p(x)$

$$V = (-3-0)^2.\frac{1}{16} + (-2-0)^2.\frac{2}{16} + (-1-0)^2.\frac{3}{16} + (0-0)^2.\frac{4}{16} + (1-0)^2.\frac{3}{16} + (2-0)^2.\frac{2}{16} + (3-0)^2.\frac{1}{16}$$

$$\mathbf{V = \frac{1}{16}(9+8+3+0+3+8+9) = \frac{40}{16} = \frac{5}{2}}$$

Standard deviation $\sigma = \sqrt{V}$

$$\boldsymbol{\sigma = \sqrt{\frac{5}{2}} = 1.58}$$

Now, $p(x \le 1) = p(-3) + p(-2) + p(-1) + p(0) + p(1)$

$$= \frac{1}{16} + \frac{2}{16} + \frac{3}{16} + \frac{4}{16} + \frac{3}{16}$$

$$\mathbf{p(x \le 1) = \frac{13}{16}}$$

$p(x > 1) = p(2) + p(3)$

$$= \frac{2}{16} + \frac{1}{16}$$

$$\mathbf{p(x > 1) = \frac{3}{16}}$$

$p(-1 < x \le 2) = p(0) + p(1) + p(2)$

$$= \frac{4}{16} + \frac{3}{16} + \frac{2}{16}$$

$$\mathbf{p(-1 < x \le 2) = \frac{9}{16}}$$

5. A random variable X has the following probability function for various values of x.

x	0	1	2	3	4	5	6	7
P(x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2 + k$

i) Find k ii) Evaluate $P(x < 6)$, $P(x \ge 6)$ and $P(3 < x \le 6)$. Also find the probability distribution and the distribution function of X.

Sol: We must have $P(x) \geq 0$ and $\sum P(x) = 1$.
The first condition is satisfied for $k \geq 0$ and we have to find k such that

$$\sum P(x) = 1$$

i.e, $0 + k + 2k + 2k + 3k + k^2 + 2k^2 + (7k^2 + k) = 1$

$$10k^2 + 9k - 1 = 0$$

$$(10k - 1)(k + 1) = 0$$

$$\mathbf{k} = \frac{\mathbf{1}}{\mathbf{10}} \quad \text{and} \quad \mathbf{k} = -\mathbf{1} \text{ (Neglect)}$$

Hence we have the following table

x	0	1	2	3	4	5	6	7
P(x)	0	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{100}$	$\frac{1}{50}$	$\frac{17}{100}$

Now, $P(x < 6) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$

$$= 0 + \frac{1}{10} + \frac{1}{5} + \frac{1}{5} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100}$$

$$\mathbf{P(x < 6) = 0.81}$$

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

$$P(x \geq 6) = \frac{1}{50} + \frac{17}{100} = \frac{19}{100}$$

$$\mathbf{P(x \geq 6) = 0.19}$$

$P(3 < x \leq 6) = P(4) + P(5) + P(6)$

$$= \frac{3}{10} + \frac{1}{100} + \frac{1}{50} = \frac{33}{100}$$

$$\mathbf{P(3 < x \leq 6) = 0.33}$$

The probability distribution is as follows

x	0	1	2	3	4	5	6	7
P(x)	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17

The distribution function of X is $f(x) = P(X \leq x) = \sum_{i=1}^{n} p(x_i)$ is called cumulative distribution function and same is as follows

x	0	1	2	3	4	5	6	7
f(x)	0	0+0.1 =0.1	0.1+0.2 =0.3	0.3+0.2 =0.5	0.5+0.3 =0.8	0.8+0.01 =0.81	0.81+0.02=0.83	0.83+0.17=1

6. A random variable X take the values $-3, -2, -1, 0, 1, 2, 3$ such hat $P(X = 0) = P(X < 0)$ and $P(X = -3) = P(X = -2) = P(X = -1) = P(X = 1) = P(X = 2) = P(X = 3)$ .Find the probability distribution.

Sol: Let the distribution be as follows

X	-3	-2	-1	0	1	2	3
P(X)	p_1	p_2	p_3	p_4	p_5	p_6	p_7

By data $P(X = 0) = P(X < 0)$

$\Rightarrow$ $P(X = 0) = P(X = -1) + P(X = -2) + P(X = -3)$

i.e., $p_4 = p_3 + p_2 + p_1$ --- (1)

Also we have

$$P(X = -3) = P(X = -2) = P(X = -1) = P(X = 1) = P(X = 2) = P(X = 3)$$

i.e., $p_1 = p_2 = p_3 = p_5 = p_6 = p_7$ --- (2)

Further we must have
$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 = 1$ --- (3)

Using (2) in (1), we get $p_4 = p_1 + p_1 + p_1$

$\mathbf{p_4 = 3p_1}$

Using (2) in (3), we get $p_1 + p_1 + p_1 + p_4 + p_1 + p_1 + p_1 = 1$

$6p_1 + p_4 = 1$

$6p_1 + 3p_1 = 1$ (since $p_4 = 3p_1$)

$9p_1 = 1$

$\mathbf{p_1 = \frac{1}{9}}$

Hence $\mathbf{p_4 = \frac{3}{9}}$

Thus , $p_2 = p_3 = p_5 = p_6 = p_7 = \frac{1}{9}$

Thus the probability distribution is as follows

X	-3	-2	-1	0	1	2	3
P(X)	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

7. The p.d.f. of variate X is given by the following table.

x	0	1	2	3	4	5	6
P(x)	K	3K	5K	7K	9K	11K	13K

For what value of K, this represents a valid probability distribution? Also find $P(x \geq 5)$ and $P(3 < x \leq 6)$.

Sol: The probability distribution is valid if $P(x) \geq 0$ and $\sum P(x) = 1$

Hence we must have $K \geq 0$ and $K + 3K + 5K + 7K + 9K + 11K + 13K = 1$

$$49K = 1\mathbf{K} = \frac{\mathbf{1}}{\mathbf{49}}$$

Also, $P(x \geq 5) = P(5) + P(6)$

$$P(x \geq 5) = 11K + 13K = 24K$$

$$\mathbf{P(x \geq 5) = \frac{24}{49}}$$

Now, $P(3 < x \leq 6) = P(4) + P(5) + P(6)$

$$P(3 < x \leq 6) = 9K + 11K + 13K$$

$$P(3 < x \leq 6) = 33K$$

$$\mathbf{P(3 < x \leq 6) = \frac{33}{49}}$$

Binomial Distribution

If **p** is the probability of success and **q** is the probability of failure, the probability of **x** success out of **n** trials is given by $\mathbf{P(x) = {}^{n}C_{x}p^{x}q^{n-x}}$

Mean and Standard deviation of the Binomial Distribution

$$\mathbf{Mean,\ \mu = np}$$

$$\mathbf{Variance,\ V = npq}$$

$$\mathbf{Std.\ Deviation,\ \sigma = \sqrt{npq}}$$

Problems

1. Find the binomial probability which has mean 2 and variance $\frac{4}{3}$.

Sol: Given Mean, $\mu = 2$ and Variance, $V = \frac{4}{3}$

For binomial distribution, $\mu = np$

$\mathrm{np} = 2$

$\mathrm{npq} = \mathrm{V}$

$2\mathrm{q} = \frac{4}{3}$

$\mathbf{q} = \frac{\mathbf{2}}{\mathbf{3}}$

Since $\mathrm{p} = 1 - \mathrm{q} = 1 - \frac{2}{3}$

$\mathbf{p} = \frac{\mathbf{1}}{\mathbf{3}}$

Since $\mathrm{np} = 2$

$\frac{n}{3} = 2$

$\mathbf{n = 6}$

The binomial probability function is $\mathbf{P(x) = {}^{n}C_{x}p^{x}q^{n-x}}$ becomes

$$\mathbf{P(x) = {}^{6}C_{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{6-x}}$$

The distribution of probability is as follows

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**
P(x)	$\left(\frac{2}{3}\right)^{6}$	${}^{6}C_{1}\left(\frac{1}{3}\right)^{1}\left(\frac{2}{3}\right)^{5}$	${}^{6}C_{2}\left(\frac{1}{3}\right)^{2}\left(\frac{2}{3}\right)^{4}$	${}^{6}C_{3}\left(\frac{1}{3}\right)^{3}\left(\frac{2}{3}\right)^{3}$	${}^{6}C_{4}\left(\frac{1}{3}\right)^{4}\left(\frac{2}{3}\right)^{2}$	${}^{6}C_{5}\left(\frac{1}{3}\right)^{5}\left(\frac{2}{3}\right)^{1}$	$\left(\frac{1}{3}\right)^{6}$

2.When a coin is tossed 4 times, find the probability of getting i) exactly one head ii) atmost 3 heads iii) atleast two heads.

Sol: Given $\mathbf{n = 4}$, $\mathbf{p = P(H)} = \frac{1}{2}$, $\mathbf{q = P(T)} = \frac{1}{2}$ (since $\mathrm{p + q = 1}$)

From binomial distribution ,the probability of x successes out of n trials is given by

$$\mathbf{P(x) = {}^{n}C_{x}p^{x}q^{n-x}}$$

$$\mathrm{P(x)} = {}^{4}C_{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{4-x}$$

$$\mathrm{P(x)} = {}^{4}C_{x}\left(\frac{1}{2}\right)^{4}$$

$$\mathbf{P(x)} = \frac{\mathbf{1}}{\mathbf{16}}.\ \mathbf{{}^{4}C_{x}}$$

i) $P(1\text{ head}) = P(x=1) = \frac{1}{16}.\ {}^4C_1 = \frac{1}{16}(4)$

$\mathbf{P(x=1) = \frac{1}{4}}$

$\mathbf{P(x=1) = 0.25}$

ii) $P(\text{atmost 3 heads}) = P(x \le 3)$

$$P(x \le 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$$

$$P(x \le 3) = \frac{1}{16}.\ {}^4C_0 + \frac{1}{16}.\ {}^4C_1 + \frac{1}{16}.\ {}^4C_2 + \frac{1}{16}.\ {}^4C_3$$

$$P(x \le 3) = \frac{1}{16}[1+4+6+4]$$

$$P(x \le 3) = \frac{15}{16}$$

$$\mathbf{P(x \le 3) = 0.9375}$$

iii) $P(\text{atleast 2 heads}) = P(x \ge 2)$

$$P(x \ge 2) = 1 - P(x < 3)$$

$$P(x \ge 2) = 1 - [P(x=0) + P(x=1)]$$

$$P(x \ge 2) = 1 - \left[\frac{1}{16}.\ {}^4C_0 + \frac{1}{16}.\ {}^4C_1\right]$$

$$P(x \ge 2) = 1 - \frac{1}{16}[1+4]$$

$$P(x \ge 2) = 1 - \frac{5}{16}$$

$$P(x \ge 2) = \frac{11}{16}$$

$$\mathbf{P(x \ge 2) = 0.6875}$$

3. In a consignment of electric lamps 5% are defective. If a random sample of 8 lamps are inspected. What is the probability that one or more lamps are defective?

Sol: Given n = 8, Probability of defective lamp $= \mathbf{p = 5\% = \frac{5}{100} = 0.05}$

$\therefore q = 1 - p = 1 - 0.05$

$\mathbf{q = 0.95}$

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x}$$

$$P(x) = 8c_x (0.05)^x (0.95)^{8-x}$$

Probability that one or more lamps are defective

$$P(x \ge 1) = 1 - P(x < 1)$$

$$P(x \ge 1) = 1 - P(x = 0)$$

$$P(x \geq 1) = 1 - 8c_0(0.05)^0(0.95)^{8-0}$$

$$\mathbf{P(x \geq 1) = 0.3366.}$$

Thus the required probability is 0.3366

4. The probability that a person aged 60 years will live upto 70 is 0.65. What is the probability that out of 10 person aged 60 atleast 7 of them will live upto 70.

Sol: Given n = 10, $\mathbf{p = 0.65}$

$\therefore q = 1 - p = 1 - 0.65$
$\mathbf{q = 0.35}$

Let x denote the number of person aged 60 years living upto 70 years.

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x}$$

$$P(x) = 10c_x(0.65)^x(0.35)^{10-x}$$

Probability that out of 10 person aged 60 atleast 7 of them will live upto 70

$P(x \geq 7) = P(7) + P(8) + P(9) + P(10)$

$P(x \geq 7) = 10c_7(0.65)^7(0.35)^3 + 10c_8(0.65)^8(0.35)^2 + 10c_9(0.65)^9(0.35)^1 + 10c_{10}(0.65)^{10}(0.35)^0$

$\mathbf{P(x \geq 7) = 0.5138}$

5.The number of telephone lines busy at an instant of time is a binomial variate with probability 0.1 that a line is busy. If 10 such lines are chosen at random, what is the probability that i) no line is busy ii) all lines are busy iii) atleast one line is busy iv) atmost 2 lines are busy.

Sol: Given n = 10, $\mathbf{p = 0.1}$

$\therefore q = 1 - p = 1 - 0.1$
$\mathbf{q = 0.9}$

Let x denote the number of telephone lines busy.

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x}$$

$\mathbf{P(x) = 10c_x(0.1)^x(0.9)^{10-x}}$

i) Probability that no line is busy

$P(x = 0) = 10c_0(0.1)^0(0.9)^{10-0} = (0.9)^{10} = \mathbf{0.3487}$

ii) Probability that all lines are busy
$P(x = 10) = 10c_{10}(0.1)^{10}(0.9)^0 = \mathbf{(0.1)^{10}}$

iii) Probability that atleast one line is busy

$P(x \geq 1) = 1 - P(x < 1)$

$P(x \geq 1) = 1 - P(x = 0)$

$P(x \geq 1) = 1 - 10c_0(0.1)^0(0.9)^{10}$

$P(x \geq 1) = 1 - 0.3487$

$\mathbf{P(x \geq 1) = 0.6513}$

iv) Probability that atmost 2 lines are busy

$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$P(x \leq 2) = (0.9)^{10} + 10c_1(0.1)^1(0.9)^2 + 10c_2(0.1)^2(0.9)^8$

$\mathbf{P(x \leq 2) = 0.9298}$

6. In a quiz contest of answering 'Yes' or 'No' what is the probability of guessing atleast 6 answers correctly out of 10 questions asked ? Also find the probability of the same if there are 4 options for a correct answer.

Sol: Given n = 10

Let x denote the correct answer .

i) $\mathbf{p = \frac{1}{2} = 0.5}$

$\therefore q = 1 - p = 1 - 0.5$

$\mathbf{q = 0.5}$

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x}$$

$$P(x) = 10c_x(0.5)^x(0.5)^{10-x}$$

$$\mathbf{P(x) = 10c_x(0.5)^{10}}$$

Probability of guessing atleast 6 answers correctly out of 10 questions

$P(x \geq 6) = P(x = 6)+P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10)$

$P(x \geq 6) = 10c_6(0.5)^{10} + 10c_7(0.5)^{10} + 10c_8(0.5)^{10} + 10c_9(0.5)^{10} + 10c_{10}(0.5)^{10}$

$P(x \geq 6) = (0.5)^{10}[10c_6 + 10c_7 + 10c_8 + 10c_9 + 10c_{10}]$

$P(x \geq 6) = (0.5)^{10}[210 + 120 + 45 + 10 + 1]$

$\mathbf{P(x \geq 6) = 0.3770}$

ii) $\mathbf{p = \frac{1}{4}}$

$\therefore q = 1 - p = 1 - \frac{1}{4}$

$\mathbf{q = \frac{3}{4}}$

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x}$$

$$P(x) = 10c_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}$$

$$P(x) = 10c_x \frac{3^{10-x}}{4^{x+10-x}}$$

$$P(x) = 10c_x \frac{3^{10-x}}{4^{10}}$$

$$\mathbf{P(x) = \frac{1}{4^{10}}[3^{10-x}.10c_x]}$$

Probability of guessing atleast 6 answers correctly out of 10 questions

$P(x \geq 6) = P(x = 6) + P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10)$

$P(x \geq 6) = \frac{1}{4^{10}}[3^4.10c_6 + 3^3.10c_7 + 3^2.10c_8 + 3.10c_9 + 10c_{10}]$

$P(x \geq 6) = \frac{1}{4^{10}}[81.210 + 27.120 + 9.45 + 3.10 + 1]$

$\mathbf{P(x \geq 6) = 0.0197}$

7. In sampling a large number of parts manufactured by a company, the mean number of defectives in samples of 20 is 2. Out of 1000 such samples how many would be expected to contain atleast 3 defective parts.

Sol: Given Mean, $\mu = 2$ and $n = 20$

$$np = 2$$

$$20p = 2$$

$$\mathbf{p = \frac{1}{10} = 0.1}$$

$$q = 1 - p$$

$q = 1 - \frac{1}{10}$ $\qquad$ $\mathbf{q = \frac{9}{10} = 0.9}$

Let x denote the defective part.

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x}$$

$\mathbf{P(x) = 20c_x(0.1)^x(0.9)^{20-x}}$

Probability of atleast 3 defective parts

$P(x \geq 3) = 1 - P(x < 3)$

$P(x \geq 3) = 1 - [P(0) + P(1) + P(2)]$

$P(x \geq 3) = 1 - [20c_0(0.1)^0(0.9)^{20} + 20c_1(0.1)^1(0.9)^{19} + 20c_2(0.1)^2(0.9)^{18}]$

$P(x \geq 3) = 1 - 0.6769$

$\mathbf{P(x \geq 3) = 0.3231}$

Thus, the number of defectives in 1000 such samples= 1000*0.3231=323.1≅**323**

8. If the mean and standard deviation of the number of correctly answered questions in a test given to 4096 students are 2.5 and $\sqrt{1.875}$. Find an estimate the number of candidates answering correctly i) 5 questions ii) 2 or less iii) 8 or more.

Sol: Given Mean = 2.5 $\sigma = \sqrt{1.875}$
Wkt, $\mu = np$ and $\sigma = \sqrt{V}$

Thus, $np = 2.5$ and $V = npq = 1.875$

Now, $\frac{npq}{np} = \frac{1.875}{2.5}$

We get q= 0.75, and p = 1-q = 0.25

n=10

Let x denote the number of correctly answered questions

$P(x) = nc_x p^x q^{n-x} = 10c_x(0.25)^x(0.75)^{10-x}$

i) Probability that 5 questions are answered correctly
$P(x = 5) = 10c_5(0.25)^5(0.75)^{10-5}$

$$\mathbf{P(x = 5) = 0.0584}$$

Estimation for 4096 students is =4096*0.0584 ≅ $\mathbf{239}$
Number of students correctly answering 5 questions is **239**

ii) Probability that 2 or less questions are answered correctly

$$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

$$= 10c_0(0.25)^0(0.75)^{10} + 10c_1(0.25)^1(0.75)^9 + 10c_2(0.25)^2(0.75)^8$$

$$\mathbf{P(x \leq 2) = 0.5256}$$

Estimation for 4096 students is =4096*0.5256 ≅ $\mathbf{2153}$

Number of students correctly answering 2 or less than 2 questions is **2153**

iii) Probability that 8 or more questions are answered correctly

$$P(x \geq 8) = P(x = 8) + P(x = 9) + P(x = 10)$$
$$= 10c_8(0.25)^8(0.75)^2 + 10c_9(0.25)^9(0.75)^1 + 10c_{10}(0.25)^{10}(0.75)^0$$
$$\mathbf{P(x \geq 8) = 0.0005}$$

Estimation for 4096 students is =4096*0.0005 $\cong$**2**

Number of students correctly answering 8 or more than 8 questions is **2**

9. In 800 families with 5 children each how many families would be expected to have i) 3 boys ii)5 girls iii) either 2 or 3 boys iv) atmost 2 girls by assuming probabilities for boys and girls to be equal.

10. An airline knows that 5% of the people making reservations on a certain flight will not turn up. Consequently, their policy is to sell 52 tickets for a flight that can only hold 50 people. What is the probability that there will be a seat for every passenger who turns up?

11. 4 coins are tossed 100 times and the following results were obtained. Fit a binomial distribution for the data and calculate the theoretical frequencies.

Number of heads	0	1	2	3	4
Frequency	5	29	36	25	5

Sol: Let x denote the number of heads and f the corresponding frequency. Since the data is in the form of a frequency distribution we shall first calculate the mean.

Mean, $\mu = \frac{\sum fx}{\sum f} = \frac{(0)(5)+(1)(29)+(2)(36)+(3)(25)+(4)(5)}{5+29+36+25+5} = \frac{196}{100} = 1.96$

From binomial distribution, $\mu = np$. Here n=4

Hence $4p = 1.96$

$$\mathbf{p = 0.49}$$

$$\mathbf{q = 0.51}$$

From binomial distribution,

$$P(x) = nc_x p^x q^{n-x} \mathbf{P(x) = 4c_x(0.49)^x(0.51)^{4-x}}$$

Since 4 coins were tossed , expected(theoretical) frequencies are obtained from

$$F(x) = 100P(x)$$

$F(x) = 100.4c_x(0.49)^x(0.51)^{4-x}$ where $x = 0,1,2,3,4.$

$F(0) = 100.4c_0(0.49)^0(0.51)^4 = 6.765 \cong 7$

$$F(1) = 100.4c_1(0.49)^1(0.51)^3 = 25.999 \cong 26$$

$$F(2) = 100.4c_2(0.49)^2(0.51)^2 = 37.47 \cong 37$$

$$F(3) = 100.4c_3(0.49)^3(0.51)^1 = 24.0004 \cong 24$$

$$F(4) = 100.4c_4(0.49)^4(0.51)^0 = 5.765 \cong 6$$

Thus the required theoretical frequencies are **7,26,37,24,6**.

Poisson Distribution

Poisson distribution is regarded as the limiting form of the binomial distribution when **n** is very large ($n \to \infty$) and **p** the probability of success is very small ($p \to 0$) so that **np** tends to a fixed finite constant say **m**.

Let m be a positive constant. Consider the infinite discrete probability distribution of a random variable x is given by $\mathbf{P(x) = \frac{m^x}{x!} e^{-m}}$

The above probability density function is called the **poisson probability** function.

Mean and Standard deviation of the Poisson Distribution

$$\textbf{Mean, } \boldsymbol{\mu} = \mathbf{m}$$

$$\textbf{Variance, } \mathbf{V} = \mathbf{m}$$

$$\textbf{Std. Deviation, } \boldsymbol{\sigma} = \sqrt{\mathbf{m}}$$

Problems

1. The probabilities of a poisson variate taking the values 3 and 4 are equal. Find the PDF.

   Also, calculate the probabilities of the variate taking the values 0 and 1.

Sol: Given $P(x = 3) = P(x = 4)$

$$\frac{m^3}{3!} e^{-m} = \frac{m^4}{4!} e^{-m}$$

$$\frac{4!}{3!} = \frac{m^4}{m^3}$$

$\therefore$ $\mathbf{m = 4.}$

From poisson distribution,

$$P(x) = \frac{m^x}{x!} e^{-m} = \frac{4^x}{x!} e^{-4}$$

$$P(x = 0) = \frac{4^0}{0!} e^{-4} = \mathbf{0.0183} \qquad P(x = 1) = \frac{4^1}{1!} e^{-4} = \mathbf{0.0733}$$

2. In a certain factory turning out razor blades there is a small probability of $\frac{1}{500}$ for any blade to be defective. The blades are supplied in packets of 10. Use poisson distribution to calculate the approximate number of packets containing i) no defective ii) one defective iii) two defective blades in a consignment of 10,000 packets.

Sol: Probability of a defective blade= $p = \frac{1}{500} = 0.002$

In a packet of 10, the mean number of defective blades is

Mean, $m = np$

$$m = 10(0.002)$$

$$\mathbf{m = 0.02}$$

From poisson distribution,

$$P(x) = \frac{m^x}{x!} e^{-m}$$

$$P(x) = \frac{(0.02)^x}{x!} e^{-(0.02)}$$

$$P(x) = (0.9802) \frac{(0.02)^x}{x!} \qquad (\text{Since } e^{-0.02} = 0.9802)$$

i) Probability of no defective

$$P(x = 0) = (0.9802) \frac{(0.02)^0}{0!} = 0.9802$$

For consignment of 10,000 packets= $\mathbf{10,000 * 0.9802 = 9802}$

ii) Probability of one defective

$$P(x = 1) = (0.9802) \frac{(0.02)^1}{1!} = 0.0196$$

For consignment of 10,000 packets= $\mathbf{10,000 * 0.0196 = 196}$

iii) Probability of one defective

$$P(x = 2) = (0.9802) \frac{(0.02)^2}{2!} = 0.0002$$

For consignment of 10,000 packets= $\mathbf{10,000 * 0.0002}$ **=2**

3. The number of accidents in a year to taxi drivers in a city follows a Poisson distribution with mean 3. Out of 1000 taxi drivers find approximately the number of the drivers with i) no accident in a year ii) more than 3 accidents in a year.

Sol: Given Mean=3

Mean, $\mu = m$

$$\mathbf{m = 3}$$

From poisson distribution,

$$P(x) = \frac{m^x}{x!} e^{-m}$$

$$P(x) = \frac{(3)^x}{x!} e^{-3}$$

$$P(x) = (0.05) \frac{(3)^x}{x!} \qquad (\text{Since } e^{-3} = 0.05)$$

i) Probability of number of drivers with no accident in a year

$$P(x = 0) = (0.05) \frac{(3)^0}{0!} = 0.05$$

Number of drivers out of 1000 with no accident in a year $\mathbf{= 1,000 * 0.05 = 50}$

ii) Probability of more than 3 accidents in a year

$$P(x > 3) = 1 - P(x \leq 3)$$

$$P(x > 3) = 1 - [P(0) + P(1) + P(2) + P(3)]$$

$$P(x > 3) = 1 - \left[(0.05)\frac{(3)^0}{0!} + (0.05)\frac{(3)^1}{1!} + (0.05)\frac{(3)^2}{2!} + (0.05)\frac{(3)^3}{3!}\right]$$

$$P(x > 3) = 1 - (0.05)[1 + 3 + 4.5 + 4.5]$$

$$P(x > 3) = 0.35$$

Number of drivers out of 1000 with more than 3 accidents in a year $= \mathbf{1,000 * 0.35 = 350}$

4. Given that 2% of the fuses manufactured by a firm are defective, find, by using poisson distribution, the probability that a box containing 200 fuses has i) no defective fuse ii) at least one defective fuse iii) exactly 3 defective fuses iv) 3 or more defective fuses.

Sol: The probability that a fuses is defective p = 2/100=0.02.

And n=200

Mean, $\mu = np = 200 * 0.02$; $\mathbf{m = 4}$

From poisson probability function $P(x) = \frac{m^x}{x!}e^{-m}$

$$P(x) = \frac{4^x}{x!}e^{-4}$$

Let x denotes the number of defective fuses in a box containing 200 fuses

i) Probability that the box contains no defective fuse

$$P(x = 0) = \frac{4^0}{0!}e^{-4} = \mathbf{0.0183}$$

ii) Probability that the box contains at least one defective fuse is

$$P(x \geq 1) = 1 - P(x < 1)$$
$$= 1 - P(x = 0)$$
$$= 1 - 0.013$$
$$\mathbf{P(x \geq 1) = 0.9817}$$

iii) Probability that the box contains exactly 3 defective fuse

$$P(x = 3) = \frac{4^3}{3!}e^{-4}$$

$$\mathbf{P(x = 3) = 0.1952}$$

iv) Probability that the box contains 3 or more defective fuses

$$P(x \geq 3) = 1 - P(x < 3)$$

$$P(x \geq 3) = 1 - \left[\frac{4^0}{0!}e^{-4} + \frac{4^1}{1!}e^{-4} + \frac{4^2}{2!}e^{-4}\right]$$

$$\mathbf{P(x \geq 3) = 0.7621}$$

5. A certain screw making machine has a chance of producing 2 defectives out of 1000. The screws are packed in boxes of 100. Using poisson distribution , find the approximate number of boxes containing i) no defective screw ii)one defective screw iii) two defective screw, in a consignment of 5000 boxes.

Sol: The probability that a screw is defective is given as p=2/1000 = 0.002.

Here n=100

Mean = np=100X0.002=0.2 ; $\mathbf{m = 0.2}$

From poisson probability function $\mathrm{P(x)} = \frac{\mathrm{m^x}}{\mathrm{x!}} \mathrm{e^{-m}}$

$$\mathrm{P(x)} = \frac{(0.2)^{\mathrm{x}}}{\mathrm{x!}} \mathrm{e^{-0.2}}$$

Let x denotes the number of defective fuses in a box containing 100 screws

i) Probability that the box contains no defective screw

$$\mathbf{P(x = 0) = \frac{0.2^0}{0!} e^{-0.2} = 0.8187}$$

Number of boxes containing no defective screw in 5000 boxes =5000*0.8187=4094

ii) Probability that the box contains one defective screw is

$$\mathrm{P(x = 1)} = \frac{(0.2)^1}{1!} \mathrm{e^{-0.2}}$$

$$= 0.2\mathrm{X}0.8187$$

$$\mathbf{P(x = 1) = 0.1637}$$

Number of boxes containing one defective screw in 5000 boxes =5000*0.1637=819

iii) Probability that the box contains two defective screw is

$$\mathrm{P(x = 2)} = \frac{(0.2)^2}{2!} \mathrm{e^{-0.2}}$$

$$= 0.02\mathrm{X}0.8187$$

$$\mathbf{P(x = 2) = 0.0163}$$

Number of boxes containing two defective screw in 5000 boxes =5000*0.0163=82

6. A communication channel receives independent pulses at the rate of 12 pulses per micro second. The probability of transmission error is 0.001 for each micro second. Compute the probabilities of i) no error during a micro second ii) one error per micro second iii) atleast one per micro second iv) two errors v) atmost two errors.

Sol: Given $\mathrm{n = 12}$ and $\mathrm{p = 0.001}$

Mean number of errors in one micro second

$$\mu = np$$

$m = 12 * 0.001$ (Since $\mu = m$, from poisson distribution)

$$\mathbf{m = 0.012}$$

From Poisson distribution

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

$$P(x) = \frac{(0.012)^x}{x!} . e^{-(0.012)}$$

$P(x) = (0.9880) \frac{(0.012)^x}{x!}$ (Since $e^{-(0.012)} = 0.9880$)

i) Probability of no error during a micro second

$$P(x = 0) = (0.9880) \frac{(0.012)^0}{0!}$$

$$\mathbf{P(x = 0) = 0.9880}$$

ii) Probability of one error during a micro second

$$P(x = 1) = (0.9880) \frac{(0.012)^1}{1!}$$

$$\mathbf{P(x = 1) = 0.0119}$$

iii) Probability of atleast one error during a micro second

$$P(x \geq 1) = 1 - P(x < 1)$$

$$P(x \geq 1) = 1 - P(x = 0)$$

$$P(x \geq 1) = 1 - P(x = 0)$$

$$P(x \geq 1) = 1 - 0.9880$$

$$\mathbf{P(x \geq 1) = 0.012}$$

iv) Probability of two error during a micro second

$$P(x = 2) = (0.9880) \frac{(0.012)^2}{2!}$$

$$\mathbf{P(x = 2) = 0.00007}$$

v) Probability of atmost two error during a micro second

$$\mathbf{P}(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

$$P(x \leq 2) = 0.9880 + 0.0119 + 0.00007$$

$$\mathbf{P(x \leq 2) = 0.99997}$$

Continuous Probability Distribution

If for every x belonging to the range of a continuous random variable X, we assign a real number f(x) satisfying the conditions,

i) $f(x) \geq 0$ ii) $\int_{-\infty}^{\infty} f(x)dx = 1$, then f(x) is called a Continuous probability function or probability density function (p.d.f)

Cumulative distribution function

If X is a continuous random variable with probability density function f(x) then the function F(x) defined by $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$ is called the Cumulative distribution function (c.d.f) of X.

NOTE: If r is any real number, then

1. $P(x \geq r) = \int_{r}^{\infty} f(x)dx$

$$P(x < r) = 1 - P(x \geq r) = 1 - \int_{r}^{\infty} f(x)dx$$

Mean and Variance

If X is a continuous random variable with probability density function f(x) where

$-\infty < x < \infty$, the mean (μ) or expectation E(X) and the variance (σ^2) of X is defined as follows.

Mean, $\boldsymbol{\mu = \int_{-\infty}^{\infty} x.\ f(x)dx}$

Variance, $\boldsymbol{\sigma^2 = \int_{-\infty}^{\infty}(x - \mu)^2 . f(x)\ dx}$ **or** $\boldsymbol{V = \int_{-\infty}^{\infty} x^2 . f(x)\ dx - \mu^2}$

Problems

1. Find which of the following is a probability density function

a) $f_1(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$ b) $f_2(x) = \begin{cases} |x|, & 0 < |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$

c) $f_3(x) = \begin{cases} 2x & , 0 < x \leq 1 \\ 4 - 4x & , 1 < x < 2 \\ 0 & , \text{otherwise} \end{cases}$

Sol: Conditions for p.d.f are $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$

a) Clearly $f_1(x) \geq 0$

$$\int_{-\infty}^{\infty} f_1(x)dx = \int_{0}^{1} f_1(x)\ dx = \int_{0}^{1} 2x\ dx = 2.\left[\frac{x^2}{2}\right]_0^1 = [1 - 0] = 1$$

$\therefore$ $f_1(x)$ is a p.d.f

b) Evidently $f_2(x) = |x| \geq 0$

$\int_{-\infty}^{\infty} f_2(x)dx = \int_{-1}^{1} f_2(x)\, dx = \int_{-1}^{1} |x|\, dx$

But, $|x| = \begin{cases} -x, & -1 < x < 0 \\ +x, & 0 < x < 1 \end{cases}$

$\therefore \int_{-\infty}^{\infty} f_2(x)dx = \int_{-1}^{0} (-x)\, dx + \int_{0}^{1} x\, dx$

$$\int_{-\infty}^{\infty} f_2(x)dx = -\left[\frac{x^2}{2}\right]_{-1}^{0} + \left[\frac{x^2}{2}\right]_{0}^{1}$$

$$\int_{-\infty}^{\infty} f_2(x)dx = -\left[0 - \frac{1}{2}\right] + \left[\frac{1}{2} - 0\right] = \frac{1}{2} + \frac{1}{2} = 1$$

$\therefore$ $f_2(x)$ is a p.d.f

c) Given $f_3(x) = 2x > 0$ in $0 < x \le 1$

But $f_3(x) = 4 - 4x$ is negative in $1 < x < 2$

The first condition is not satisfied.

$\therefore$ $f_3(x)$ is not a p.d.f

2. Find the value of c such that $f(x) = \begin{cases} \frac{x}{6} + c, & 0 \le x \le 3 \\ 0, & \text{elsewhere} \end{cases}$ is a p.d.f . Also find $P(1 \le x \le 2)$.

Sol: Here $f(x) \ge 0$ if $c \ge 0$; Also we must have $\int_{-\infty}^{\infty} f(x)dx = 1$

i.e., $\int_{0}^{3} \left(\frac{x}{6} + c\right) dx = 1$

i.e., $\left[\frac{x^2}{12} + cx\right]_{0}^{3} = 1$

i.e., $\frac{1}{12}(9 - 0) + c(3 - 0) = 1$

i.e., $\frac{9}{12} + 3c = 1$

i.e., $3c = 1 - \frac{9}{12}$

$3c = \frac{3}{12}$

$\therefore \mathbf{c = \frac{1}{12}}$

Now, $P(1 \le x \le 2) = \int_{1}^{2} f(x)dx$

$$= \int_{1}^{2} \left(\frac{x}{6} + \frac{1}{12}\right) dx$$

$$= \left[\frac{x^2}{12} + \frac{x}{12}\right]_1^2$$

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

$$= \frac{3}{12} + \frac{1}{12} = \frac{4}{12}$$

$$\mathbf{P(1 \le x \le 2) = \frac{1}{3}}$$

3. Find the constant k such that $f(x) = \begin{cases} kx^2, & 0 < x < 3 \\ 0, & \text{otherwise} \end{cases}$ is a p.d.f . Also compute

i) $P(1 < x < 2)$ ii) $P(x \le 1)$ iii) $P(x > 1)$ iv) Mean v) Variance

Sol: : Here $f(x) \ge 0$ if $k \ge 0$; Also we must have $\int_{-\infty}^{\infty} f(x)dx = 1$

i.e., $\int_0^3 kx^2\, dx = 1$

i.e., $\left[\frac{kx^3}{3}\right]_0^3 = 1$

i.e., $\frac{k}{3}(27-0) = 1$

i.e., $9k = 1$

i.e., $\mathbf{k = \frac{1}{9}}$

$$\therefore f(x) = \begin{cases} \frac{x^2}{9}, & 0 < x < 3 \\ 0, & \text{otherwise} \end{cases}$$

i) $P(1 < x < 2) = \int_1^2 f(x)dx$

$$= \int_1^2 \frac{x^2}{9}\, dx$$

$$= \left[\frac{x^3}{27}\right]_1^2$$

$$= \frac{1}{27}(8-1)$$

$$\mathbf{P(1 < x < 2) = \frac{7}{27}}$$

ii) $P(x \le 1) = \int_0^1 f(x)dx$

$$P(x \le 1) = \int_0^1 \frac{x^2}{9}\, dx$$

$$= \left[\frac{x^3}{27}\right]_0^1$$

$$= \frac{1}{27}(1-0)$$

$$\mathbf{P(x \leq 2) = \frac{1}{27}}$$

iii) $P(x > 1) = \int_1^3 f(x)dx$

$$= \int_1^3 \frac{x^2}{9}\ dx$$

$$P(x > 1) = \left[\frac{x^3}{27}\right]_1^3$$

$$= \frac{1}{27}(27 - 1)$$

$$\mathbf{P(x > 1) = \frac{26}{27}}$$

iv) Mean, $\mu = \int_{-\infty}^{\infty} x.\ f(x)dx$

$$= \int_0^3 x.\frac{x^2}{9}dx$$

$$= \int_0^3 \frac{x^3}{9}dx$$

$$\mu = \left[\frac{x^4}{36}\right]_0^3$$

$$\mu = \frac{1}{36}(81 - 0)$$

$$\mu = \frac{81}{36}$$

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

v)Variance, $V = \int_{-\infty}^{\infty} x^2.f(x)\ dx - \mu^2$

$$= \int_0^3 x^2.\frac{x^2}{9}dx - \left(\frac{9}{4}\right)^2$$

$$= \int_0^3 \frac{x^4}{9}dx - \frac{81}{16}$$

$$= \left[\frac{x^5}{45}\right]_0^3 - \frac{81}{16}$$

$$V = \frac{1}{45}[3^5 - 0] - \frac{81}{16}$$

$$V = \frac{243}{45} - \frac{81}{16}$$

$$V = \frac{81}{240}$$

$$\mathbf{V = \frac{27}{80}}$$

Exponential Distribution

The continuous probability distribution having the probability density function f(x) given by

$f(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$, where $\alpha > 0$ is known as the **exponential distribution**.

Evidently $f(x) \geq 0$ and we have $\int_{-\infty}^{\infty} f(x)dx = 1$.

f(x) satisfy both the conditions required for a continuous probability function.

Mean and standard deviation of the Exponential Distribution

Mean , $\boldsymbol{\mu = \frac{1}{\alpha}}$

Std. deviation , $\boldsymbol{\sigma = \frac{1}{\alpha}}$

Problems

1. If x is an exponential variate with mean 3 find i) $P(x > 1)$ ii) $P(x < 3)$

Sol: The p.d.f of the exponential distribution is given by

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{for } 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

The mean of the distribution is given by $\frac{1}{\alpha}$.

By data, $\mu = \frac{1}{\alpha} = 3$

$$\therefore \ \boldsymbol{\alpha = \frac{1}{3}}$$

Hence , $f(x) = \begin{cases} \frac{1}{3} e^{-\frac{x}{3}}, & \text{for } 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$

$$\text{i) } P(x > 1) = 1 - P(x \leq 1)$$

$$P(x > 1) = 1 - P(x \leq 1)$$

$$= 1 - \int_0^1 f(x)dx$$

$$= 1 - \int_0^1 \frac{1}{3} e^{-\frac{x}{3}} dx$$

$$= 1 - \frac{1}{3}\int_0^1 e^{-\frac{x}{3}} dx$$

$$= 1 - \frac{1}{3}\left[\frac{e^{-\frac{x}{3}}}{-\frac{1}{3}}\right]_0^1$$

$$= 1 + \left[e^{-\frac{1}{3}} - e^{-0}\right]$$

$$= 1 + e^{-\frac{1}{3}} - 1$$

$$= e^{-\frac{1}{3}}$$

$\mathbf{P(x > 1) = 0.7165}$

ii) $P(x < 3) = \displaystyle\int_0^3 f(x)dx$

$$= \int_0^3 \frac{1}{3} e^{-\frac{x}{3}} dx$$

$$= \frac{1}{3} \int_0^3 e^{-\frac{x}{3}} dx$$

$$= \frac{1}{3} \left[\frac{e^{-\frac{x}{3}}}{-\frac{1}{3}} \right]_0^3$$

$$= -[e^{-1} - e^{-0}]$$

$$= -e^{-1} + 1 = 1 - \frac{1}{e}$$

$\mathbf{P(x < 3) = 0.6321}$

2. The length of telephone conversation in a booth has been an exponential distribution and found on an average to be 5 minutes. Find the probability that a random call made from this booth i) ends less than 5 minutes ii) between 5 and 10 minutes.

Sol: We have, $f(x) = \alpha e^{-\alpha x}, \quad 0 < x < \infty$

By data, $\mu = \frac{1}{\alpha} = 5$

$$\therefore \quad \boldsymbol{\alpha = \frac{1}{5}}$$

$f(x) = \dfrac{1}{5} e^{-\frac{x}{5}}, \quad 0 < x < \infty$

i) $P(x < 5) = \displaystyle\int_0^5 f(x)dx$

$$= \int_0^5 \frac{1}{5} e^{-\frac{x}{5}} dx$$

$$= \frac{1}{5} \int_0^5 e^{-\frac{x}{5}} dx$$

$P(x < 5) = \dfrac{1}{5} \left[\dfrac{e^{-\frac{x}{5}}}{-\frac{1}{5}} \right]_0^5$

$$= -[e^{-1} - e^{-0}]$$

$$= -e^{-1} + 1 = 1 - \frac{1}{e}$$

$$\mathbf{P(x < 5) = 0.6321}$$

ii) $P(5 < x < 10) = \int_{5}^{10} f(x)dx$

$$= \int_{5}^{10} \frac{1}{5} e^{-\frac{x}{5}} dx$$

$$= \frac{1}{5} \int_{5}^{10} e^{-\frac{x}{5}} dx$$

$$P(5 < x < 10) = \frac{1}{5} \left[\frac{e^{-\frac{x}{5}}}{-\frac{1}{5}} \right]_{5}^{10}$$

$$= -[e^{-2} - e^{-1}]$$

$$= -e^{-2} + e^{-1} = \frac{1}{e} - \frac{1}{e^2}$$

$\mathbf{P(5 < x < 10) = 0.2325}$

3. In a certain town the duration of a shower is exponentially distributed with mean 5 minutes. What is the probability that a shower will last for: (i) 10 minutes or more (ii) less than 10 minutes (iii) between 10 and 12 minutes.

Sol: We have, $f(x) = \alpha e^{-\alpha x}, \quad 0 < x < \infty$

By data, $\mu = \frac{1}{\alpha} = 5$

$\therefore \quad \boldsymbol{\alpha = \frac{1}{5}}$

$$f(x) = \frac{1}{5} e^{-\frac{x}{5}}, \quad 0 < x < \infty$$

i) $P(x \geq 10) = \int_{10}^{\infty} f(x)dx$

$$= \int_{10}^{\infty} \frac{1}{5} e^{-\frac{x}{5}} dx$$

$$= \frac{1}{5} \int_{10}^{\infty} e^{-\frac{x}{5}} dx$$

$$= \frac{1}{5} \left[\frac{e^{-\frac{x}{5}}}{-\frac{1}{5}} \right]_{10}^{\infty}$$

$$= -[e^{-\infty} - e^{-2}]$$

$$= -e^{-\infty} + e^{-2} = -0 + e^{-2}$$

$$\mathbf{P(x \geq 10) = 0.1353}$$

ii) $P(x < 10) = 1 - P(x \geq 10)$

$P(x < 10) = 1 - 0.1353$

$\mathbf{P(x < 10) = 0.8647}$

iii) $P(10 < x < 12) = \int_{10}^{12} f(x)dx$

$$= \int_{10}^{12} \frac{1}{5} e^{-\frac{x}{5}} dx$$

$$= \frac{1}{5} \int_{10}^{12} e^{-\frac{x}{5}} dx$$

$$= \frac{1}{5} \left[\frac{e^{-\frac{x}{5}}}{-\frac{1}{5}} \right]_{10}^{12}$$

$$= -[e^{-2.4} - e^{-2}]$$

$P(10 < x < 12) = e^{-2} - e^{-2.4}$

$\mathbf{P(10 < x < 12) = 0.0446}$

4.The life of a compressor manufactured by a company is known to be 200 months on an average following an exponential distribution. Find the probability that the life of a compressor of that company is i) less than 200 months ii) between 100 months and 25 years.

Sol: We have, $f(x) = \alpha e^{-\alpha x}, \quad 0 < x < \infty$

By data, $\mu = \frac{1}{\alpha} = 200$

$\therefore \ \boldsymbol{\alpha = \frac{1}{200}}$

$$f(x) = \frac{1}{200} e^{-\frac{x}{200}}, \quad 0 < x < \infty$$

i) $P(x < 200) = \int_{0}^{200} f(x)dx$

$$= \int_{0}^{200} \frac{1}{200} e^{-\frac{x}{200}} dx$$

$$= \frac{1}{200} \int_{0}^{200} e^{-\frac{x}{200}} dx$$

$$= \frac{1}{200} \left[\frac{e^{-\frac{x}{200}}}{-\frac{1}{200}} \right]_{0}^{200}$$

$$= -[e^{-1} - e^{-0}]$$

$$= -e^{-1} + 1 = 1 - \frac{1}{e}$$

$$\mathbf{P(x < 200) = 0.6321}$$

Now, 25 years=25*12 months=300 months

ii) $P(100 < x < 300) = \int_{100}^{300} f(x)dx$

$$= \int_{100}^{300} \frac{1}{200} e^{-\frac{x}{200}} dx$$

$$= \frac{1}{200} \int_{100}^{300} e^{-\frac{x}{200}} dx$$

$$= \frac{1}{200} \left[\frac{e^{-\frac{x}{200}}}{-\frac{1}{200}} \right]_{100}^{300}$$

$$= -[e^{-1.5} - e^{-2}]$$

$$\mathbf{P(100 < x < 300) = 0.3834}$$

5. Determine k if random variable x has the density function $f(x) = \frac{k}{1+x^2}, \; -\infty < x < \infty$ hence find i) $P(x \geq 0)$ ii) $P(0 < x < 1)$ iii) c.d.f

Sol: Given f(x) is a probability density function.

$\therefore \; f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$

i.e., $\int_{-\infty}^{\infty} \frac{k}{1+x^2} dx = 1$

$$2k \int_0^{\infty} \frac{1}{1+x^2} dx = 1$$

$$2k[\tan^{-1} x]_0^{\infty} = 1 \qquad ; \qquad 2k[\tan^{-1}(\infty) - \tan^{-1}(0)] = 1$$

$$2k\left[\frac{\pi}{2} - 0\right] = 1$$

$$k.\pi = 1 \mathbf{k} = \frac{\mathbf{1}}{\boldsymbol{\pi}}$$

$$f(x) = \frac{\frac{1}{\pi}}{1 + x^2}$$

i) $P(x \geq 0) = \int_0^{\infty} f(x)dx$

$$= \int_0^{\infty} \frac{\frac{1}{\pi}}{1+x^2} dx$$

$$= \frac{1}{\pi}\int_0^{\infty} \frac{1}{1+x^2}\,dx$$

$$= \frac{1}{\pi}\left[\tan^{-1}x\right]_0^{\infty}$$

$$P(x \geq 0) = \frac{1}{\pi}\left[\tan^{-1}(\infty) - \tan^{-1}(0)\right]$$

$$P(x \geq 0) = \frac{1}{\pi}\left[\frac{\pi}{2} - 0\right]$$

$$\mathbf{P(x \geq 0) = \frac{1}{2}}$$

$$ii)P(0 < x < 1) = \int_0^1 f(x)\,dx$$

$$= \int_0^1 \frac{\frac{1}{\pi}}{1+x^2}\,dx$$

$$= \frac{1}{\pi}\int_0^1 \frac{1}{1+x^2}\,dx$$

$$P(0 < x < 1) = \frac{1}{\pi}\left[\tan^{-1}x\right]_0^1$$

$$= \frac{1}{\pi}\left[\tan^{-1}(1) - \tan^{-1}(0)\right]$$

$$= \frac{1}{\pi}\left[\frac{\pi}{4} - 0\right]$$

$$\mathbf{P(0 < x < 1) = \frac{1}{4}}$$

$$iii)c.d.f = \int_{-\infty}^{x} f(x)\,dx$$

$$= \int_{-\infty}^{x} \frac{\frac{1}{\pi}}{1+x^2}\,dx$$

$$= \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+x^2}\,dx$$

$$c.d.f = \frac{1}{\pi}\left[\tan^{-1}x\right]_{-\infty}^{x}$$

$$= \frac{1}{\pi}\left[\tan^{-1}(x) - \tan^{-1}(-\infty)\right]$$

$$= \frac{1}{\pi}\left[\tan^{-1}x - \left(-\frac{\pi}{2}\right)\right]$$

$$\mathbf{c.d.f = \frac{1}{\pi}\left[\tan^{-1}x + \frac{\pi}{2}\right] \quad if \quad -\infty < x < \infty.}$$

6. The daily turn over in a medical shop is exponentially distributed with Rs.6000 as the average with a net profit of 8%. Find the probability that the net profit exceeds Rs.500 on a randomly chosen day.

Sol: Let ‘x’ denote the turn over per day.

∴ By exponential distribution

$$f(x) = \alpha e^{-\alpha x} \ ; x > 0.$$

Given, By data, $\mu = \frac{1}{\alpha} = 6000$

$$\therefore \ \boldsymbol{\alpha} = \frac{\mathbf{1}}{\mathbf{6000}}$$

$$f(x) = \frac{1}{6000} e^{-\frac{x}{6000}}, \quad 0 < x < \infty$$

Let ‘A’ be the turn over for which the net profit is Rs.500.

Since the net profit is 8% of the turn over,

$$A * \frac{8}{100} = 500$$

$$\Rightarrow \ \boldsymbol{A = 6250}$$

∴ The probability of profit exceeding Rs.500 is

$$P(x > 6250) = \int_{6250}^{\infty} f(x)dx$$

$$P(x > 6250) = \int_{6250}^{\infty} \frac{1}{6000} e^{-\frac{x}{6000}} dx$$

$$= \frac{1}{6000} \int_{6250}^{\infty} e^{-\frac{x}{6000}} dx$$

$$= \frac{1}{6000} \left[\frac{e^{-\frac{x}{6000}}}{-\frac{1}{6000}} \right]_{6250}^{\infty}$$

$$= -\left[e^{-\infty} - e^{-\frac{6250}{6000}} \right]$$

$$= e^{-\frac{6250}{6000}}$$

$$\mathbf{P(x > 6250) = 0.3529}$$

Normal Distribution

The continuous probability distribution having the probability density function f(x) given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ where $-\infty < x < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$ is known as the normal distribution.

Evidently $f(x) \geq 0$ and we have $\int_{-\infty}^{\infty} f(x)dx = 1$.

f(x) satisfy both the conditions required for a continuous probability function.

Mean and standard deviation of the Normal Distribution

$$\mathbf{Mean = \mu}$$

$$\mathbf{Variance = \sigma^2}$$

Geometrically we can write

i) $\int_{-\infty}^{\infty} \boldsymbol{\phi(z)dz = 1}$ **ii)** $\int_{-\infty}^{0} \boldsymbol{\phi(z)dz} = \int_{0}^{\infty} \boldsymbol{\phi(z)dz} = \frac{1}{2}$

Note

1. $P(-\infty \leq z \leq \infty) = 1$
2. $P(-\infty \leq z \leq 0) = \frac{1}{2}$
3. $P(0 \leq z \leq \infty)$ **or** $P(z \geq 0) = \frac{1}{2}$

Also, $P(-\infty < z < z_1) = P(-\infty < z \leq 0) + P(0 \leq z < z_1)$

i.e., $\mathbf{P(z < z_1) = 0.5 + \phi(z_1)}$

Also, $P(z > z_2) = P(z \geq 0) - P(0 \leq z < z_2)$

i.e., $\mathbf{P(z > z_2) = 0.5 - \phi(z_2)}$

Problems

1. Evaluate the following probabilities with the help of normal probability tables

i) $P(z \geq 0.85)$ ii) $P(-1.64 \leq z \leq -0.88)$ iii) $P(z \leq -2.43)$ iv) $P(|z| \leq 1.94)$

Sol: i) $P(z \geq 0.85) = P(z \geq 0) - P(0 \leq z \leq 0.85)$

$$= 0.5 - \phi(0.85)$$

$$= 0.5 - 0.3023$$

$$\mathbf{P(z \geq 0.85) = 0.1977}$$

ii) $P(-1.64 \le z \le -0.88)$

By Symmetry, $P(-1.64 \le z \le -0.88) = P(0.88 \le z \le 1.64)$

$$= P(0 \le z \le 1.64) - P(0 \le z \le 0.88)$$

$$= \phi(1.64) - \phi(0.88)$$

$$= 0.4495 - 0.3106$$

$$\mathbf{P(-1.64 \le z \le -0.88) = 0.1389}$$

iii) $P(z \le -2.43) = P(z \ge 2.43)$

$P(z \le -2.43) = P(z \ge 0) - P(0 \le z \le 2.43)$

$$= 0.5 - \phi(2.43)$$

$$= 0.5 - 0.4925$$

$$\mathbf{P(z \ge -2.43) = 0.0075}$$

iv) $P(|z| \le 1.94) = P(-1.94 \le z \le 1.94)$

$$= 2\, P(0 \le z \le 1.94)$$

$$= 2.\phi(1.94)$$

$$= 2.(0.4738)$$

$$\mathbf{P(|z| \le 1.94) = 0.9476}$$

2. If x is a normal variate with mean 30 and standard deviation 5 find the probability that
i) $26 \le x \le 40$ ii) $x \ge 45$

Sol: We have standard normal variate, $z = \frac{x-\mu}{\sigma} = \frac{x-30}{5}$ [By data $\mu = 30,\ \sigma = 5$]

i) To find $\mathbf{P(26 \le x \le 40)}$

If $x = 26$, $z = \frac{26-30}{5} = -0.8$; If $x = 40$, $z = \frac{40-30}{5} = 2$

Hence we need to find , $\mathbf{P(-0.8 \le z \le 2)}$

$$P(-0.8 \le z \le 2) = P(-0.8 \le z \le 0) + P(0 \le z \le 2)$$

$$= P(0 \le z \le 0.8) + P(0 \le z \le 2)$$

$$= \phi(0.8) + \phi(2)$$

$$= 0.2881 + 0.4772$$

$$\mathbf{P(-0.8 \le z \le 2) = 0.7653}$$

ii) To find $\mathbf{P(x \ge 45)}$

If $x = 45$, $z = \frac{45-30}{5} = 3$

Hence we need to find , $\mathbf{P(z \ge 3)}$

$P(z \geq 3) = P(z \geq 0) - P(0 \leq z \leq 3)$

$= 0.5 - \phi(3)$

$= 0.5 - 0.4987$

$\mathbf{P(z \geq 3) = 0.0013}$

3. The marks of 1000 students in an examination follows a normal distribution with mean 70 and standard deviation 5. Find the number of students whose marks will be

i) less than 65 ii) more than 75 iii) 65 to 75

Sol: Let x represent the marks of students

We have standard normal variate, $z = \frac{x-\mu}{\sigma} = \frac{x-70}{5}$ [By data $\mu = 70,\ \sigma = 5$]

i) To find $\mathbf{P(x < 65)}$

If $x = 65$, $z = \frac{65-70}{5} = -1$

Hence we need to find , $\mathbf{P(z < -1)}$

$P(z < -1) = P(z > 1)$

$= P(z \geq 0) - P(0 \leq z \leq 1)$

$= 0.5 - \phi(1)$

$= 0.5 - 0.3413$

$\mathbf{P(z < -1) = 0.1587}$

Number of students scoring less than 65 marks=1000*0.1587= $\mathbf{158.7 \cong 159}$

ii) To find $\mathbf{P(x > 75)}$

If $x = 75$, $z = \frac{75-70}{5} = 1$

Hence we need to find , $\mathbf{P(z > 1)}$

$P(z > 1) = P(z \geq 0) - P(0 \leq z \leq 1)$

$= 0.5 - \phi(1)$

$= 0.5 - 0.3413$

$\mathbf{P(z > 1) = 0.1587}$

Number of students scoring more than 75 marks=1000*0.1587= $\mathbf{158.7 \cong 159}$

iii) To find $\mathbf{P(65 < x < 75)}$

Hence we need to find , $\mathbf{P(-1 < z < 1)}$

$\mathbf{P(-1 < z < 1) = 2P(0 < z < 1)}$

$= 2\phi(1)$

$= 2(0.3413)$

$\mathbf{P(-1 < z < 1) = 0.6826}$

Number of students scoring marks between 65 to 75 marks=1000*0.6826= $\mathbf{682.6 \cong 683}$

4. In a test on electric bulbs, it was found that the life time of a particular brand was distributed normally with an average life of 2000 hours and S.D of 60 hours. If a firm purchases 2500 bulbs, find the number of bulbs that are likely to last for i) more than 2100 hours ii) less than 1950 hours iii) between 1900 to 2100 hours.

Sol: Given $\mu = 2000, \quad \sigma = 60$

We have standard normal variate, $\mathbf{z = \frac{x-\mu}{\sigma} = \frac{x-2000}{60}}$

i) To find $\mathbf{P(x > 2100)}$

If $x = 2100$, $z = \frac{2100-2000}{60} = 1.67$

Hence we need to find , $\mathbf{P(z > 1.67)}$

$P(z > 1.67) = P(z \geq 0) - P(0 < z < 1.67)$

$= 0.5 - \phi(1.67)$

$= 0.5 - 0.4525$

$\mathbf{P(z > 1.67) = 0.0475}$

∴ **Number of bulbs that are likely to last for more than 2100 hours =** $\mathbf{2500 * 0.0475}$

$\mathbf{= 118.75 \cong 119}$

ii) To find $\mathbf{P(x < 1950)}$

If $x = 1950$, $z = \frac{1950-2000}{60} = -0.83$

Hence we need to find , $\mathbf{P(z < -0.83)}$

$P(z < -0.83) = P(z > 0.83)$

$= P(z \geq 0) - P(0 < z < 0.83)$

$= 0.5 - \phi(0.83)$

$$= 0.5 - 0.2967$$

$$\mathbf{P(z < -0.83) = 0.2033}$$

∴ **Number of bulbs that are likely to last for more than 1950 hours** $= \mathbf{2500 * 0.2033}$

$$= \mathbf{508.25 \cong 508}$$

iii) To find $\mathbf{P(1900 < x < 2100)}$

If $x = 1900$, $z = \frac{1900-2000}{60} = -1.67$; If $x = 2100$, $z = \frac{2100-2000}{60} = 1.67$

Hence we need to find , $\mathbf{P(-1.67 < z < 1.67)}$

$$P(-1.67 < z < 1.67) = 2P(0 < z < 1.67)$$

$$= 2 * \phi(1.67)$$

$$= 2 * 0.4525$$

$$\mathbf{P(-1.67 < z < 1.67) = 0.905}$$

∴ **Number of bulbs that are likely to last between1900 and 2100 hours**

$$= \mathbf{2500 * 0.2033}$$

$$= \mathbf{2262.5 \cong 2263}$$

5. In a normal distribution 31% of the items are under 45 and 8% of the items are over 64. Find the mean and S.D of the distribution.

Sol: Let μ and σ be the mean and S.D of the normal distribution.

By data, $P(x < 45) = 0.31$ and $P(x > 64) = 0.08$

We have standard normal variate, $\mathbf{z = \frac{x-\mu}{\sigma}}$

When $x = 45$, $z = \frac{45-\mu}{\sigma} = z_1$ (Say)

$x = 64$, $z = \frac{64-\mu}{\sigma} = z_2$ (Say)

So we have ,

$P(z < z_1) = 0.31$ and $P(z > z_2) = 0.08$

i.e., $0.5 + \varphi(z_1) = 0.31$ and $0.5 - \varphi(z_2) = 0.08$

$\varphi(z_1) = 0.31 - 0.5$ and $\varphi(z_2) = 0.5 - 0.08$

$\varphi(z_1) = -0.19$ and $\varphi(z_2) = 0.42$

Referring to the normal probability tables, we have

$0.1915(\cong 0.19) = \varphi(0.5)$ and $0.4192(\cong 0.42) = \varphi(1.4)$

$\varphi(z_1) = -\varphi(0.5)$ and $\varphi(z_2) = \varphi(1.4)$

$$\Rightarrow \quad \mathbf{z_1 = -0.5} \quad \text{and} \quad \mathbf{z_2 = 1.4}$$

i.e., $\frac{45-\mu}{\sigma} = -0.5$ and $\frac{64-\mu}{\sigma} = 1.4$

$45 - \mu = -0.5\sigma$ and $64 - \mu = 1.4\sigma$

$\mu - 0.5\sigma = 45$ and $\mu + 1.4\sigma = 64$

Solving above equations, we get

$$\boldsymbol{\mu = 50} \quad \text{and} \quad \boldsymbol{\sigma = 10}$$

Thus, Mean=50 and S.D=10

6. The mean weight of 500 students during a medical examination was found to be 50 kgs and S.D weight 6 kgs. Assuming that the weights are normally distributed, find the number of student having weight i) between 40 and 50 kgs ii) more than 60 kgs. $[\varphi(1.67) = 0.4525]$

Sol: Given $\mu = 50, \quad \sigma = 6$

We have standard normal variate, $\mathbf{z} = \frac{\mathbf{x-\mu}}{\boldsymbol{\sigma}} = \frac{\mathbf{x-50}}{\mathbf{6}}$

i) **To find $\mathbf{P(40 < x < 50)}$**

If $x = 40$, $z = \frac{40-50}{6} = -1.67$; If $x = 50$, $z = \frac{50-50}{6} = 0$

Hence we need to find , $\mathbf{P(-1.67 < z < 0)}$

$P(-1.67 < z < 0) = P(0 < z < 1.67)$

$= \phi(1.67)$

$\mathbf{P(-1.67 < z < 0)} = 0.4525$

∴ **Number of students having weight between 40 and 50 kgs= $\mathbf{500 * 0.4525 = 226.25}$**

$\mathbf{\cong 226}$

ii) To find $\mathbf{P(x > 60)}$

If $x = 60$, $z = \frac{60-50}{6} = 1.67$

Hence we need to find , $\mathbf{P(z > 1.67)}$

$P(z > 1.67) = P(z \geq 0) - P(0 < z < 1.67)$

$= 0.5 - \phi(1.67)$

$= 0.5 - 0.4525$

$\mathbf{P(z > 1.67) = 0.0475}$

∴ **Number of students having weight more than 60 kgs = $\mathbf{500 * 0.0475}$**

$\mathbf{= 23.75 \cong 24}$

7. In an examination 7% of students score less than 35% marks and 89% of students score less than 60% of marks. Find the mean and S.D of the distribution if the marks are normally distributed. It is given that P(1.2263)=0.39 and P(1.4757)=0.43.

Sol: Let μ and σ be the mean and S.D of the normal distribution.

By data, $P(x < 35) = 0.07$ and $P(x < 60) = 0.89$

We have standard normal variate, $\mathbf{z} = \frac{\mathbf{x-\mu}}{\boldsymbol{\sigma}}$

When $x = 35$, $z = \frac{35-\mu}{\sigma} = z_1$ (Say)

$x = 60$, $z = \frac{60-\mu}{\sigma} = z_2$ (Say)

So we have ,

$$P(z < z_1) = 0.07 \quad \text{and} \quad P(z < z_2) = 0.89$$

i.e., $0.5 + \varphi(z_1) = 0.07$ and $0.5 + \varphi(z_2) = 0.89$

$\varphi(z_1) = 0.07 - 0.5$ and $\varphi(z_2) = 0.89 - 0.5$

$\varphi(z_1) = -0.43$ and $\varphi(z_2) = 0.39$

Using the given data in the RHS of these, we have

$$\varphi(z_1) = -\varphi(1.4757) \quad \text{and} \quad \varphi(z_2) = \varphi(1.2263)$$

$$\Rightarrow \quad \mathbf{z_1 = -1.4757} \quad \text{and} \quad \mathbf{z_2 = 1.2263}$$

i.e., $\frac{35-\mu}{\sigma} = -1.4757$ and $\frac{60-\mu}{\sigma} = 1.2263$

$$35 - \mu = -1.4757\sigma \quad \text{and} \quad 60 - \mu = 1.2263\sigma$$

$$\mu - 1.4757\sigma = 35 \quad \text{and} \quad \mu + 1.2263\sigma = 60$$

Solving above equations, we get

$$\boldsymbol{\mu = 48.65} \quad \text{and} \quad \boldsymbol{\sigma = 9.25}$$

Thus, Mean = 48.65 and S.D = 9.25

MODULE-2

JOINT PROBABILITY DISTRIBUTIONS AND MARKOV CHAIN

Joint Probability Distribution

If X and Y are two discrete random variables, we define the *joint probability function* of X and Y by $\mathrm{P}(X = x, Y = y) = f(x, y)$ where $f(x, y)$ satisfy the conditions $f(x, y) \geq 0$ and $\sum_x \sum_y f(x, y) = 1$.

Suppose $X = \{x_1, x_2, \ldots., x_m\}$ and $Y = \{y_1, y_2, \ldots., y_n\}$ then,

$\mathrm{P}\left(X = x_i, Y = y_j\right) = f(x_i, y_j)$ is denoted by J_{ij} .

Expectation, Variance, Covariance and Correlation

If X is a discrete random variable taking values $x_1, x_2, \ldots., x_n$ having probability function $f(x)$ then,

The ***expectation*** of X is denoted by $\mu_X = E(X) = \sum_{i=1}^{n} x_i f(x_i)$

The ***Variance*** of X is denoted by $V(X) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$, where μ is the mean of X.

$$V = E(X^2) - \mu_X^2$$

The ***Standard deviation*** of X is denoted by $\sigma_X = \sqrt{V(X)}$

If X and Y are two discrete random variables having the joint probability distribution $\mathrm{f(x, y)}$ then the expectation of X and Y are defined as

$$\mu_X = E(X) = \sum_i x_i f(x_i)$$

$$\mu_Y = E(Y) = \sum_j y_j g(y_j)$$

$$and \qquad E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

If X and Y are random variables having mean $\mu_X[\mathrm{or} E(X)]$ and $\mu_Y[\mathrm{or} E(Y)]$ respectively, then the covariance of X and Y are defined as

$$\mathrm{Cov(X, Y) = E(XY) - E(X)E(Y)}$$

The *correlation* of X and Y denoted by $\rho(X, Y)$ is given by $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ where

$\sigma_X^2 = E(X^2) - \mu_X^2, \quad \sigma_Y^2 = E(Y^2) - \mu_Y^2$

Note: If X and Y are independent random variables then

i) $E(XY) = E(X)E(Y)$

ii) $Cov(X, Y) = 0 \;\; and \; \rho(X, Y) = 0$

Problems

1. The joint distribution of two random variables X and Y is as follows

$Y \rightarrow$ $X \downarrow$	-4	2	7
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
5	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

Compute the following

i) $E(X)$ and $E(Y)$ ii) $E(XY)$ iii) σ_X and σ_Y iv) $Cov(X,Y)$ v) $\rho(X,Y)$

Sol: The distribution (marginal distribution) of X and Y is as follows.

$Y \rightarrow$ $X \downarrow$	- 4	2	7	$f(x_i)$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}+\frac{1}{4}+\frac{1}{8}=\frac{1}{2}$
5	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2}$
$g(y_j)$	$\frac{1}{8}+\frac{1}{4}=\frac{3}{8}$	$\frac{1}{4}+\frac{1}{8}=\frac{3}{8}$	$\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$	1

Distribution of X

x_i	1	5
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

Distribution of Y

y_j	- 4	2	7
$g(y_j)$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$

i) $E(X) = \sum_i x_i f(x_i)$ $\qquad$ $E(Y) = \sum_j y_j g(y_j)$

$E(X) = (1)\left(\frac{1}{2}\right) + (5)\left(\frac{1}{2}\right)$ $\qquad$ $E(Y) = (-4)\left(\frac{3}{8}\right) + (2)\left(\frac{3}{8}\right) + (7)\left(\frac{1}{4}\right)$

$E(X) = \frac{1}{2} + \frac{5}{2}$ $\qquad$ $E(Y) = -\frac{3}{2} + \frac{3}{4} + \frac{7}{4}$

Thus, $E(X) = \mu_X = 3$ $\qquad$ $E(Y) = \mu_Y = 1$

ii) $E(XY) = \sum_{i,j} x_i y_j J_{ij}$

$E(XY) = (1)(-4)\left(\frac{1}{8}\right) + (1)(2)\left(\frac{1}{4}\right) + (1)(7)\left(\frac{1}{8}\right) + (5)(-4)\left(\frac{1}{4}\right) + (5)(2)\left(\frac{1}{8}\right) + (5)(7)\left(\frac{1}{8}\right)$

$E(XY) = -\frac{1}{4} + \frac{1}{2} + \frac{7}{8} - 5 + \frac{5}{4} + \frac{35}{8}$

Thus, $E(XY) = \frac{3}{2}$

iii) $\sigma_X^2 = E(X^2) - \mu_X^2$; $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

Now, $E(X^2) = \sum_i x_i^2 f(x_i)$ $\qquad$ $E(Y^2) = \sum_j y_j^2 f(y_j)$

$E(X^2) = (1^2)\left(\frac{1}{2}\right) + (5^2)\left(\frac{1}{2}\right)$ $\qquad$ $E(Y^2) = ((-4)^2)\left(\frac{3}{8}\right) + (2^2)\left(\frac{3}{8}\right) + (7^2)\left(\frac{1}{4}\right)$

$E(X^2) = 13$ $\qquad$ $E(Y^2) = \frac{79}{4}$

$\therefore\ \sigma_X^2 = 13 - 3^2$; $\sigma_Y^2 = \frac{79}{4} - 1^2$

$\sigma_X^2 = 4$; $\sigma_Y^2 = \frac{75}{4}$

$\sigma_X = 2$; $\sigma_Y = 4.33$

iv) $Cov(X,Y) = E(XY) - E(X)E(Y)$

$Cov(X,Y) = \frac{3}{2} - (3)(1)$

$Cov(X,Y) = -\frac{3}{2}$

v) $\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

$\rho(X,Y) = \frac{-\frac{3}{2}}{(2)(4.33)}$

$\rho(X,Y) = -0.1732$

2. The joint probability distribution table for two random variables X and Y is as follows

$Y \rightarrow$ $X \downarrow$	-2	-1	4	5
1	0.1	0.2	0	0.3
2	0.2	0.1	0.1	0

Determine the marginal distributions of X and Y . Also compute i) Expectations of X, Y and XY ii) Standard deviations of X, Y iii) Covariance of X and Y iv) correlation of X and Y. Further verify that X and Y are dependent random variables. Also find $P(X + Y > 0)$.

Sol: Marginal distribution of X

x_i	1	2
$f(x_i)$	0.6	0.4

Marginal distribution of Y

y_j	-2	-1	4	5
$g(y_j)$	0.3	0.3	0.1	0.3

i) $E(X) = \sum_i x_i f(x_i)$ $\qquad$ $E(Y) = \sum_j y_j g(y_j)$

$E(X) = (1)(0.6) + (2)(0.4)$ $\qquad$ $E(Y) = (-2)(0.3) + (-1)(0.3) + (4)(0.1) + (5)(0.3)$

$E(X) = 0.6 + 0.8$ $\qquad$ $E(Y) = -0.6 - 0.3 + 0.4 + 1.5$

Thus, $E(X) = \mu_X = 1.4$ $\qquad$ $E(Y) = \mu_Y = 1$

$$E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$$E(XY) = (1)(-2)(0.1) + (1)(-1)(0.2) + (1)(4)(0) + (1)(5)(0.3) + (2)(-2)(0.2) + (2)(-1)(0.1) + (2)(4)(0.1)$$

$+(2)(5)(0)$

$$E(XY) = -0.2 - 0.2 + 0 + 1.5 - 0.8 - 0.2 + 0.8 + 0$$

Thus, $E(XY) = 0.9$

ii) $\sigma_X^2 = E(X^2) - \mu_X^2$ $\quad ; \quad$ $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

Now, $E(X^2) = \sum_i x_i^2 f(x_i)$ $\qquad$ $E(Y^2) = \sum_j y_j^2 g(y_j)$

$E(X^2) = (1^2)(0.6) + (2^2)(0.4)$ $\qquad$ $E(Y^2) = ((-2)^2)(0.3) + ((-1)^2)(0.3) + (4^2)(0.1) + (5^2)(0.3)$

$E(X^2) = 2.2 \qquad\qquad E(Y^2) = 10.6$

$\therefore\ \sigma_X^2 = 2.2 - (1.4)^2 \quad ; \quad \sigma_Y^2 = 10.6 - 1^2$

$\sigma_X^2 = 0.24 \quad ; \quad \sigma_Y^2 = 9.6$

$\sigma_X = 0.49 \quad ; \quad \sigma_Y = 3.1$

iii) $Cov(X, Y) = E(XY) - E(X)E(Y)$

$Cov(X, Y) = 0.9 - (1.4)(1)$

$Cov(X, Y) = -0.5$

iv) $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

$\rho(X, Y) = \frac{-0.5}{(0.49)(3.1)}$

$\rho(X, Y) = -0.3$

If X and Y are independent random variables we must have $f(x_i).\,g(y_j) = J_{ij}$

It can be seen that $f(x_1).\,g(y_1) = (0.6)(0.3) = 0.18$ and $J_{11} = 0.1$ i.e., $f(x_i).\,g(y_j) \neq J_{ij}$

Similarly for others also the condition is not satisfied.

Hence we conclude that X and Y are dependent random variables.

We have, $X = \{x_i\} = \{x_1, x_2\} = \{1,2\}$ respectively.

$Y = \{y_j\} = \{y_1, y_2, y_3, y_4\} = \{-2, -1, 4, 5\}$ respectively.

Also, $J_{11} = 0.1$, $J_{12} = 0.2$, $J_{13} = 0$, $J_{14} = 0.3$

$J_{21} = 0.2$, $J_{22} = 0.1$, $J_{23} = 0.1$, $J_{24} = 0$

$X + Y > 0$ is possible when (X, Y) take the values

$(x_1, y_3) = (1,4)$; $(x_1, y_4) = (1,5)$; $(x_2, y_2) = (2, -1)$; $(x_2, y_3) = (2,4)$ and

$(x_2, y_4) = (2,5)$

Hence , $P(X + Y > 0) = J_{13} + J_{14} + J_{22} + J_{23} + J_{24}$

$P(X + Y > 0) = 0 + 0.3 + 0.1 + 0.1 + 0$

Thus, $P(X + Y > 0) = 0.5$

3. Suppose X and Y are independent random variables with the following respective distribution, find the joint distribution of X and Y. Also verify $Cov(X,Y) = 0$.

x_i	1	2
$f(x_i)$	0.7	0.3

y_j	-2	5	8
$g(y_j)$	0.3	0.5	0.2

Sol: Since X and Y are independent, the joint distribution $J(x,y)$ is obtained by using the definition $f(x_i).g(y_j) = J_{ij}$.

J_{ij} are obtained on multiplication of the marginal entries.

From the given data , the required J_{ij} is found

$$J_{11} = f(x_1).g(y_1) = (0.7)(0.3) = 0.21$$

$$J_{12} = f(x_1).g(y_2) = (0.7)(0.5) = 0.35$$

$$J_{13} = f(x_1).g(y_3) = (0.7)(0.2) = 0.14$$

$$J_{21} = f(x_2).g(y_1) = (0.3)(0.3) = 0.09$$

$$J_{22} = f(x_2).g(y_2) = (0.3)(0.5) = 0.15$$

$$J_{23} = f(x_2).g(y_3) = (0.3)(0.2) = 0.06$$

The joint distribution table is as follows

$Y \rightarrow$ $X \downarrow$	-2	5	8	$f(x_i)$
1	0.21	0.35	0.14	0.7
2	0.09	0.15	0.06	0.3
$g(y_j)$	0.3	0.5	0.2	1

We have $Cov(X,Y) = E(XY) - E(X)E(Y)$

$E(X) = \sum_i x_i f(x_i)$

$E(X) = (1)(0.7) + (2)(0.3)$

$E(X) = 0.7 + 0.6$

$E(Y) = \sum_j y_j g(y_j)$

$E(Y) = (-2)(0.3) + (5)(0.5) + (8)(0.2)$

$E(Y) = -0.6 + 2.5 + 1.6$

Thus, $E(X) = \mu_X = 1.3$ $\qquad$ $E(Y) = \mu_Y = 3.5$

$$E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$E(XY) = (1)(-2)(0.21) + (1)(5)(0.35) + (1)(8)(0.14) + (2)(-2)(0.09) + (2)(5)(0.15) + (2)(8)(0.06)$

$E(XY) = -0.42 + 1.75 + 1.12 - 0.36 + 1.5 + 0.96$

Thus, $E(XY) = 4.55$

Hence $Cov(X, Y) = 4.55 - (1.3)(3.5) = 0$

Thus the result $Cov(X, Y) = 0$ for independent random variables X and Y is verified.

4. X and Y are independent random variables. X take values 2,5,7 with probability $1/2, 1/4, 1/4$ respectively. Y take values 3,4,5 with the probability $1/3, 1/3, 1/3$.

i) Find the joint probability distribution of X and Y

ii) Show that the covariance of X and Y is equal to zero.

iii) Find the probability distribution of $Z = X + Y$

Sol: Given

x_i	2	5	7
$f(x_i)$	$1/2$	$1/4$	$1/4$

y_j	3	4	5
$g(y_j)$	$1/3$	$1/3$	$1/3$

i) We have $J_{ij} = f(x_i).g(y_j)$ where $i, j = 1,2,3$

$J_{11} = f(x_1).g(y_1) = 1/2 . 1/3 = 1/6$ $\qquad$ $J_{12} = f(x_1).g(y_2) = 1/2 . 1/3 = 1/6$

$J_{13} = f(x_1).g(y_3) = 1/2 . 1/3 = 1/6$ $\qquad$ $J_{21} = f(x_2).g(y_1) = 1/4 . 1/3 = 1/12$

$J_{22} = f(x_2).g(y_2) = 1/4 . 1/3 = 1/12$ $\qquad$ $J_{23} = f(x_2).g(y_3) = 1/4 . 1/3 = 1/12$

$J_{31} = f(x_3).g(y_1) = 1/4 . 1/3 = 1/12$ $\qquad$ $J_{32} = f(x_3).g(y_2) = 1/4 . 1/3 = 1/12$

$J_{33} = f(x_3).g(y_3) = 1/4 . 1/3 = 1/12$

The joint distribution table is as follows

$Y \rightarrow$ $X \downarrow$	3	4	5	$f(x_i)$
2	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_2$
5	$^1/_{12}$	$^1/_{12}$	$^1/_{12}$	$^1/_4$
7	$^1/_{12}$	$^1/_{12}$	$^1/_{12}$	$^1/_4$
$g(y_j)$	$^1/_3$	$^1/_3$	$^1/_3$	1

ii) We have $Cov(X,Y) = E(XY) - E(X)E(Y)$

$E(X) = \sum_i x_i f(x_i)$ $\qquad$ $E(Y) = \sum_j y_j g(y_j)$

$E(X) = (2)(^1/_2) + (5)(^1/_4) + (7)(^1/_4)$ $\qquad$ $E(Y) = (3)(^1/_3) + (4)(^1/_3) + (5)(^1/_3)$

$E(X) = 4$ $\qquad$ $E(Y) = 4$

Thus, $E(X) = \mu_X = 4$ $\qquad$ $E(Y) = \mu_Y = 4$

$$E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$E(XY) = (2)(3)(^1/_6) + (2)(4)(^1/_6) + (2)(5)(^1/_6) + (5)(3)(^1/_{12}) + (5)(4)(^1/_{12}) + (5)(5)(^1/_{12}) + (7)(3)(^1/_{12}) + (7)(4)(^1/_{12}) + (7)(5)(^1/_{12})$

$E(XY) = 16$

Thus, $E(XY) = 16$

Thus, $Cov(X,Y) = 16 - (4)(4) = 0$

iii) $Z = X + Y$

Let $z_i = x_i + y_i$ and hence $\{z_i\} = \{5,6,7,8,9,10,11,12\}$

The corresponding probabilities are ,

$^1/_6, {}^1/_6, {}^1/_6, {}^1/_{12}, {}^1/_{12}, {}^1/_{12} + {}^1/_{12}, {}^1/_{12}, {}^1/_{12}$

The probability distribution of $Z = X + Y$ is as follows

Z	5	6	7	8	9	10	11	12
P(Z)	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_{12}$	$^1/_{12}$	$^1/_{12} + {}^1/_{12} = {}^1/_6$	$^1/_{12}$	$^1/_{12}$

We note that $\sum P(Z) = 1$.

5. The joint probability distribution of two discrete random variables X and Y is given by $f(x,y) = k(2x + y)$ where x and y are integers such that $0 \leq x \leq 2$ $0 \leq y \leq 3$.
i) Find the value of the constant k.
ii) Find the marginal probability distributions of X and Y.
iii) Show that the random variables X and Y are dependent.
iv) Compute $E(X)E(Y), E(XY)$
v) Compute $E(X^2), E(Y^2)$
vi) Compute σ_X, σ_Y
vii) Find $P(X = 1, Y = 2)$, $P(X = 2, Y = 1)$
viii) Find $P(X \geq 1, Y \leq 2)$, $P(X + Y > 2)$

Sol: $X = \{x_i\} = \{0,1,2\}$ and $Y = \{y_j\} = \{0,1,2,3\}$

$f(x,y) = k(2x + y)$ and the joint probability distribution table is formed as follows

$Y \rightarrow$ $X \downarrow$	0	1	2	3	Sum
0	0	k	$2k$	$3k$	$6k$
1	$2k$	$3k$	$4k$	$5k$	$14k$
2	$4k$	$5k$	$6k$	$7k$	$22k$
Sum	$6k$	$9k$	$12k$	$15k$	$42k$

i) We must have , $42k = 1$

$$\boldsymbol{k = \frac{1}{42}}$$

ii) Marginal probability distribution is as follows

x_i	**0**	**1**	**2**
$f(x_i)$	$6/42$	$14/42$	$22/42$

y_j	**0**	**1**	**2**	**3**
$g(y_j)$	$6/42$	$9/42$	$12/42$	$15/42$

iii) It can be easily seen that $f(x_i).g(y_j) \neq J_{ij}$

Hence the random variables are dependent.

iv) $E(X) = \sum_i x_i f(x_i)$

$E(X) = (0)(^6/_{42}) + (1)(^{14}/_{42}) + (2)(^{22}/_{42})$

$E(X) = {^{58}/_{42}}$

Thus, $\boldsymbol{E(X) = \mu_X = {^{29}/_{21}}}$

$E(Y) = \sum_j y_j g(y_j)$

$E(Y) = (0)(^6/_{42}) + (1)(^9/_{42}) + (2)(^{12}/_{42}) + (3)(^{15}/_{42})$

$E(Y) = {^{39}/_{21}}$

$\boldsymbol{E(Y) = \mu_Y = {^{13}/_7}}$

$$E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$$E(XY) = (0)(0 + k + 4k + 9k) + (1)(0 + 3k + 8k + 15k) + (2)(0 + 5k + 12k + 21k)$$

$$E(XY) = 0 + 26k + 76k$$

$$E(XY) = 102k = \frac{102}{42}$$

Thus, $\boldsymbol{E(XY) = \frac{17}{7}}$

v) $E(X^2) = \sum_i x_i^2 f(x_i)$

$E(X^2) = (0) + (1^2)\left(\frac{14}{42}\right) + (2^2)\left(\frac{22}{42}\right)$

$E(X^2) = \frac{102}{42}$

Thus, $\boldsymbol{E(X^2) = \frac{17}{7}}$

$E(Y^2) = \sum_j y_j^2 g(y_j)$

$E(Y^2) = (0) + (1^2)\left(\frac{9}{42}\right) + (2^2)\left(\frac{12}{42}\right) + (3^2)\left(\frac{15}{42}\right)$

$E(Y^2) = \frac{192}{42}$

$\boldsymbol{E(Y^2) = \frac{32}{7}}$

vi) $\sigma_X^2 = E(X^2) - \mu_X^2$

$$\sigma_X^2 = \frac{17}{7} - \left(\frac{29}{21}\right)^2$$

$$\sigma_X^2 = \frac{230}{441}$$

$$\sigma_X = \frac{\sqrt{230}}{21}$$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2$

$$\sigma_Y^2 = \frac{32}{7} - \left(\frac{13}{7}\right)^2$$

$$\sigma_Y^2 = \frac{495}{441}$$

$$\sigma_Y = \frac{\sqrt{495}}{21}$$

Thus, $\boldsymbol{\sigma_X = 0.72}$ and $\boldsymbol{\sigma_Y = 1.06}$

vii) Let $X = \{x_i\} = \{x_1, x_2, x_3\} = \{0,1,2\}$ respectively.

$Y = \{y_j\} = \{y_1, y_2, y_3, y_4\} = \{0,1,2,3\}$ respectively.

Also, $J_{11} = 0$, $J_{12} = {^1/_{42}}$, $J_{13} = {^2/_{42}}$, $J_{14} = {^3/_{42}}$

$J_{21} = {^2/_{42}}$, $J_{22} = {^3/_{42}}$, $J_{23} = {^4/_{42}}$, $J_{24} = {^5/_{42}}$

$J_{31} = {^4/_{42}}$, $J_{32} = {^5/_{42}}$, $J_{33} = {^6/_{42}}$, $J_{34} = {^7/_{42}}$

Now, $P(X = 1, Y = 2) = f(x_2, y_3) = J_{23} = {^4/_{42}} = {^2/_{21}}$

$P(X = 2, Y = 1) = f(x_3, y_2) = J_{32} = {}^{5}/_{42}$

viii) $(X, Y) = \{(1,0), (1,1), (1,2), (2,0), (2,1), (2,2)\} = \{(x_2, y_1), (x_2, y_2), (x_2, y_3), (x_3, y_1), (x_3, y_2), (x_3, y_3)\}$

$$P(X \geq 1, Y \leq 2) = J_{21}+J_{22}+J_{23}+J_{31}+J_{32}+J_{33}$$

$$P(X \geq 1, Y \leq 2) = \frac{2}{42} + \frac{3}{42} + \frac{4}{42} + \frac{4}{42} + \frac{5}{42} + \frac{6}{42} = \frac{24}{42}$$

$$\boldsymbol{P(X \geq 1, Y \leq 2) = \frac{4}{7}}$$

To find $\boldsymbol{P(X + Y > 2)}$

$$(X, Y) = \{(\mathbf{0}, \mathbf{3})(\mathbf{1}, \mathbf{2})(\mathbf{1}, \mathbf{3})(\mathbf{2}, \mathbf{1})(\mathbf{2}, \mathbf{2})(\mathbf{2}, \mathbf{3})\}$$
$$= \{(x_1, y_4), (x_2, y_3), (x_2, y_4), (x_3, y_2), (x_3, y_3), (x_3, y_4)\}$$

$$P(X + Y > 2) = J_{14}+J_{23}+J_{24}+J_{32}+J_{33}+J_{34}$$

$$P(X + Y > 2) = \frac{3}{42} + \frac{4}{42} + \frac{5}{42} + \frac{5}{42} + \frac{6}{42} + \frac{7}{42} = \frac{30}{42}$$

$$\boldsymbol{P(X + Y > 2) = \frac{5}{7}}$$

6. A fair coin is tossed thrice. The random variables $X\ and\ Y$ are defined as follows.
$X = 0\ or\ 1$ according as head or tail occurs on the first toss. $Y = Number\ of\ heads$
i) Determine the distributions of $X\ and\ Y$.
ii) Determine the joint distribution of $X\ and\ Y$.
iii) Obtain the expectations of $X, Y\ and\ XY$. Also find S.D's of $X\ and\ Y$.
iv) Compute the covariance and correlation of $X\ and\ Y$.

Sol: The sample space S and the association of random variables $X\ and\ Y$ is given by the following table

S	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	0	0	0	0	1	1	1	1
Y	3	2	2	1	2	1	1	0

i) The probability distribution of $X\ and\ Y$ is found as follows.

$X = \{x_i\} = \{0,1\}$ and $Y = \{y_j\} = \{0,1,2,3\}$

$P(X = 0) = \frac{4}{8} = \frac{1}{2}$; $P(X = 1) = \frac{4}{8} = \frac{1}{2}$

$P(Y = 0) = \frac{1}{8}$; $P(Y = 1) = \frac{3}{8}$; $P(Y = 2) = \frac{3}{8}$; $P(Y = 3) = \frac{1}{8}$

Thus, we have the following probability distribution of $X\ and\ Y$.

x_i	**0**	**2**
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

y_j	**0**	**1**	**2**	**3**
$g(y_j)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

ii) The joint distribution of X *and* Y is found by computing $J_{ij} = P(X = x_i, Y = y_j)$ where

we have $x_1 = 0,\ x_2 = 1$ and $y_1 = 0, y_2 = 1, y_3 = 2, y_4 = 3$

$J_{11} = P(X = 0, Y = 0) = 0$ ($X = 0$ implies that there is a head turn out and Y the total number of heads is 0, outcome is impossible)

$J_{12} = P(X = 0, Y = 1) = \frac{1}{8}$ (Corresponding to the out come HTT)

$J_{13} = P(X = 0, Y = 2) = \frac{2}{8} = \frac{1}{4}$ (out comes are HHT and HTH)

$J_{14} = P(X = 0, Y = 3) = \frac{1}{8}$ (out come is HHH)

$J_{21} = P(X = 1, Y = 0) = \frac{1}{8}$; out come is TTT

$J_{22} = P(X = 1, Y = 1) = \frac{2}{8} = \frac{1}{4}$; out come is THT, TTH

$J_{23} = P(X = 1, Y = 2) = \frac{1}{8}$; out come is THH

$J_{24} = P(X = 1, Y = 3) = 0$; since the out come is impossible.

The required joint probability distribution of X *and* Y is as follows.

$Y \rightarrow$ $X \downarrow$	**0**	**1**	**2**	**3**	**Sum**
0	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{1}{2}$
Sum	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

iii) $E(X) = \sum_i x_i f(x_i)$ $\qquad E(Y) = \sum_j y_j g(y_j)$

$E(X) = (0)(1/2) + (1)(1/2)$ $\qquad E(Y) = (0)(1/8) + (1)(3/8) + (2)(3/8) + (3)(1/8)$

$E(X) = {}^{1}/_{2}$ $\qquad$ $E(Y) = {}^{12}/_{8}$

Thus, $\boldsymbol{E(X) = \mu_X = {}^{1}/_{2}}$ $\qquad$ $\boldsymbol{E(Y) = \mu_Y = {}^{3}/_{2}}$

Now, $E(XY) = \sum_{i,j} x_i y_j J_{ij}$

$E(XY) = 0 + (1)(1)({}^{1}/_{4}) + (1)(2)({}^{1}/_{8}) + 0$

$E(XY) = {}^{1}/_{4} + {}^{2}/_{8}$

Thus, $\boldsymbol{E(XY) = {}^{1}/_{2}}$

Also, $\sigma_X^2 = E(X^2) - \mu_X^2$ $\qquad$ $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

Now, $E(X^2) = \sum_i x_i^2 f(x_i)$ $\qquad$ $E(Y^2) = \sum_j y_j^2 f(y_j)$

$E(X^2) = (0^2)\left(\frac{1}{2}\right) + (1^2)\left(\frac{1}{2}\right)$ $\qquad$ $E(Y^2) = (0)\left(\frac{1}{8}\right) + (1^2)\left(\frac{3}{8}\right) + (2^2)\left(\frac{3}{8}\right) + (3^2)\left(\frac{1}{8}\right)$

$\boldsymbol{E(X^2) = \frac{1}{2}}$ $\qquad$ $\boldsymbol{E(Y^2) = 3}$

$\sigma_X^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2$ $\qquad$ $\sigma_Y^2 = 3 - \left(\frac{3}{2}\right)^2$

$\sigma_X^2 = \frac{1}{4}$ $\qquad$ $\sigma_Y^2 = \frac{3}{4}$

$\boldsymbol{\sigma_X = \frac{1}{2}}$ $\qquad$ $\boldsymbol{\sigma_Y = \frac{\sqrt{3}}{2}}$

iv) We have $Cov(X,Y) = E(XY) - E(X)E(Y)$

$$Cov(X,Y) = \frac{1}{2} - \frac{1}{2} \cdot \frac{3}{2}$$

$$\boldsymbol{Cov(X,Y) = -\frac{1}{4}}$$

Also, $\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

$$\rho(X,Y) = \frac{-\frac{1}{4}}{\left(\frac{1}{2}\right)\left(\frac{\sqrt{3}}{2}\right)}$$

$$\boldsymbol{\rho(X,Y) = -\frac{1}{\sqrt{3}}}$$

7. Let X be a random variable with the following distribution and Y is defined to be X^2

x_i	-1	-2	1	2
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Determine i) the distribution g of Y ii) the joint distribution of X and Y

iii) E(X), $E(Y), E(XY)$ iv) Cov(X, Y) v) $\rho(X, Y)$

Sol: Here Y is defined to be X^2.

The distribution g of Y is given by

y_j	**1**	**4**
$g(y_j)$	$\frac{1}{2}$	$\frac{1}{2}$

ii) The joint distribution of X and Y is

$\boldsymbol{Y} \rightarrow$ $\boldsymbol{X} \downarrow$	**1**	**4**
-2	0	$\frac{1}{4}$
-1	$\frac{1}{4}$	0
1	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$

iii) $E(X) = \sum_i x_i f(x_i)$

$E(X) = (-1)(1/4) + (-2)(1/4) + (1)(1/4) + (2)(1/4)$

$E(X) = 0$

Thus, $\boldsymbol{E(X) = \mu_X = 0}$

$E(Y) = \sum_j y_j g(y_j)$

$E(Y) = (1)(1/2) + (4)(1/2)$

$E(Y) = 5/2$

$\boldsymbol{E(Y) = \mu_Y = 5/2}$

Now, $E(XY) = \sum_{i,j} x_i y_j J_{ij}$

$E(XY) = (-2)(0) + (-2)(1) + (-1)(1/4) + (-1)(0) + (1)(1/4) + (1)(0) + (2)(0) + (2)(1)$

$E(XY) = -2 - 1/4 + 1/4 + 2$

Thus, $\boldsymbol{E(XY) = 0}$

Also, $\sigma_X^2 = E(X^2) - \mu_X^2$ $\sigma_Y^2 = E(Y^2) - \mu_Y^2$

Now, $E(X^2) = \sum_i x_i^2 f(x_i)$ $E(Y^2) = \sum_j y_j^2 f(y_j)$

$E(X^2) = ((-1)^2)\left(\frac{1}{4}\right) + ((-2)^2)\left(\frac{1}{4}\right) + ((1)^2)\left(\frac{1}{4}\right) + ((2)^2)\left(\frac{1}{4}\right)$

$\boldsymbol{E(X^2) = \frac{5}{2}}$

$E(Y^2) = (1^2)\left(\frac{1}{2}\right) + (4^2)\left(\frac{1}{2}\right)$

$\boldsymbol{E(Y^2) = \frac{17}{2}}$

iv) $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$

$Cov(X, Y) = 0 - 0\left({}^{5}/_{2}\right)$

$\boldsymbol{Cov(X, Y) = 0}$

v) $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

$\rho(X, Y) = \frac{0}{(1.581)(1.658)}$

$\boldsymbol{\rho(X, Y) = 0}$

8. The joint probability function of two discrete random variables X and Y is given by $f(x, y) = cxy$ for $x = 1,2,3$ and $y = 1,2,3$. Find the following

i) Constant c ii) $P(X = 2,\ Y = 3)$ iii) $P(1 \leq X \leq 2,\ Y \leq 2)$ iv) $P(X \geq 2)$ v) $P(Y < 2)$

vi) $P(X = 1)$ v) $P(Y = 3)$

MODULE-3&4

STATISTICAL INFERENCE-1&2

Sampling Theory

Random Sampling

A large collection individuals or attributes or numerical data can be understood as a *population* or *universe*.

A finite subset of the universe is called a *sample*. The number of individuals in a sample is called a *sample size(n)*.

If $n \leq 30$ then the sample is said to be *small* otherwise it is a *large sample*.

The selection of an individual or item from the population in such a way that each has the same chance of being selected is called as *random sampling*.

Sampling where a member of the population may be selected more than once is called as *sampling with replacement*.

If a member cannot be chosen more than once is called as *sampling without replacement*.

Parameter and Statistics

The statistical constants of population is mean(μ) , standard deviation (σ), correlation (ρ) are called the *parameter*.

The statistical constants drawn from the given population is mean$(\bar{x})$, standard deviation (s), correlation (r) are called *statistic*.

Testing of Hypothesis

To reach the decisions about the populations and the basis of sample information, we make certain assumptions about the population involved. Such assumptions which may or may not be true are called *hypothesis*. **OR** Quantitative statement about the population. Explanation on the basis of limited evidences.

Null Hypothesis

The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called the *Null Hypothesis* denoted by H_0.

Any hypothesis which is complimentary to the null hypothesis is called *Alternative Hypothesis* denoted by H_1.

Type-I and Type-II Errors

Type-I Error: Wrong decision to reject the null hypothesis when it is actually true.

Type-II Error: Wrong decision to accept the null hypothesis when it is actually false.

	Accepting the hypothesis	Rejecting the hypothesis
Hypothesis True	Correct decision	Wrong decision (Type-I Error)
Hypothesis False	Wrong decision (Type-II Error)	Correct decision

Significance level

The probability level, below which leads to the rejection of the hypothesis is known as the *significance level*. The probability levels are fixed at 0.05 or 0.01 being 5% or 1%. These are called *Significance level*.

Test of Significance

The process which enables us to decide about the acceptance or rejection of the hypothesis is called the *test of significance*.

Confidence Intervals

Let us suppose that we have a normal population with mean μ and S.D σ . If $\bar{x}$ is the sample

mean of a random sample size n the quantity z defined by $z = \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ is called the *Standard normal variate*.

From the normal distribution table we find the 95% of the area lies between $z = -1.96$ and $z = +1.96$ i.e., 95% confidence interval lies between -1.96 and $+1.96$.

Further 5% level of significance is denoted by $z_{0.05}$.Mathematically we can write

$$-1.96 \leq \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \leq 1.96$$

$$-\frac{\sigma}{\sqrt{n}}(1.96) \leq \bar{x}-\mu \leq (1.96)\frac{\sigma}{\sqrt{n}}$$

$$\mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}}(1.96) \quad \text{and} \quad \bar{x} - \frac{\sigma}{\sqrt{n}}(1.96) \leq \mu$$

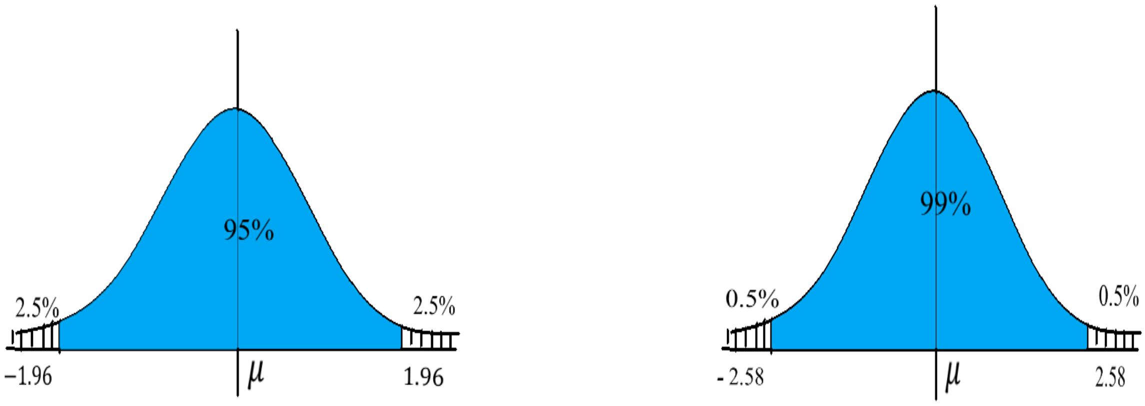
Also we can write combining the two results in the form,

$$\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{--- (1)}$$

Similarly, from the table of normal areas 99% of the area lies between -2.58 and 2.58. This is equivalent to the form,

$$\bar{x} - 2.58\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 2.58\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{--- (2)}$$

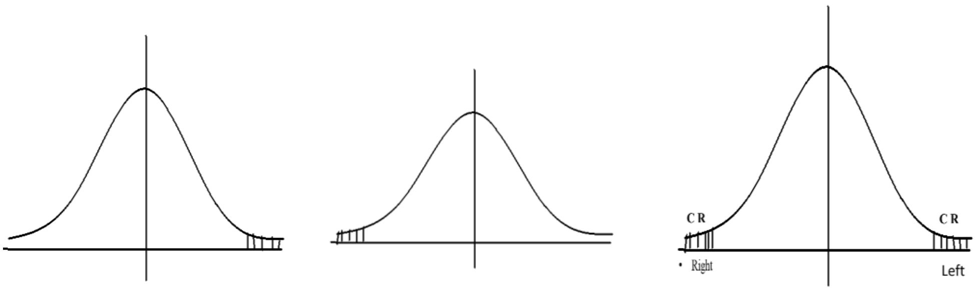
We can say that (1) is the 95% confidence interval and (2) is the 99% confidence interval.



One tailed and two tailed tests

In our test of acceptance or non acceptance of a hypothesis, we concentrated on the value of z on both sides of the mean. This can be categorically stated that, the focus of the attention lies in the two “tails” of the distribution and hence such a test is called a *two tailed test*.

Sometimes we will be interested in the extreme values to only one side of the mean in which the region of significance will be a region to one side of the distribution. Obviously the area of such a region will be equal to the level of significance itself. Such a tail is called a *one tailed test*.



One tailed test **Two tailed test**

The following table will be useful for working problems

Test	Critical Value of z	
	5% level	**1% level**
One-tailed test	-1.645 or 1.645	-2.33 or 2.33

Two-tailed test	-1.96 or 1.96	-2.58 or 2.58

Tests of significance of Proportions

1. Let x be the observed number of successes in a sample size of n and $\mu = np$ be the expected number of successes let the associated standard normal variate Z be defined by $z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$

2. If $|Z|$ > 2.58, we conclude that the differences is highly significant and reject the hypothesis .Probable limits $= p \pm 2.58\sqrt{\frac{pq}{n}}$

3. If Z < 1.96, difference between the observed and expected number of successes is not significant.

4. If Z >1.96, difference is significant at 5% level of significance.

Test of hypothesis for means

Let μ_1 and μ_2 be the mean of two populations. Let $(\bar{x}_1, \sigma_1)$; $(\bar{x}_2, \sigma_2)$ be the mean and standard deviation of two large samples of size n_1 and n_2 respectively. To test the null hypothesis H_0 that there is no difference between the population means. i.e., $H_0: \mu_1 = \mu_2$

The static for this is given by $\boldsymbol{Z} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Test of significance of proportions

1.A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin in unbiased.

Sol: Let us suppose that the coin is unbiased

p = Probability of getting a head in one toss = ½

Since *p* + *q* = *1*, *q*= ½

Expected number of heads in 1000 tosses = *np*

= 1000 × ½ = 500

Actual Number of heads = 540 = x

Now $x - np$ = 540 − 500 = 40

Consider $z = \frac{x-\mu}{\sqrt{npq}} = \frac{x-np}{\sqrt{npq}} = \frac{540-500}{\sqrt{1000X0.5X0.5}}$

z =2.53<2.58

Thus, we can say that the coin is unbiased.

2. A sample of 900 days was taken in a coastal town and it was found that on 100 days the weather was very hot. Obtain the probable limits of the percentage of very hot weather.

Sol: Probability of very hot weather , $p = \frac{100}{900} = \frac{1}{9}$

$\therefore \; q = \frac{8}{9} \quad , n = 900$

Probable limits $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

$= \frac{1}{9} \pm (2.58)\sqrt{\frac{1}{9} \cdot \frac{8}{9} \cdot \frac{1}{900}}$

$= 0.111 \pm 0.027$

$= 0.084 \;\; and \;\; 0.138$

Probability limits of very hot weather is 8.4% to 13.8%

3. In a sample of 500 men it was found that 60% of them had over weight. What can we infer about the proportion of people having over weight in the population?

Sol: Probability of persons having over weight is , $p = \frac{60}{100} = 0.6$

$\therefore \; q = 1 - p = 0.4 \quad , n = 500$

Probable limits $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

$= 0.6 \pm (2.58)\sqrt{\frac{(0.6)(0.4)}{500}}$

$= 0.6 \pm 0.0565$

Probable limits $= 0.\ 5435 \;\; and \;\; 0.6565$

Thus, the probable limits of people having over weight is 54.35% and 65.65%

4. A survey was conducted in a slum locality of 2000 families by selecting a sample size 800. It was revealed that 180 families were illiterates. Find the probable limits of the illiterate families in the population of 2000.

Sol: Probability of illiterate families is , $p = \frac{180}{800} = 0.225$

$\therefore \; q = 1 - p = 0.775 \quad , n = 800$

Probable limits $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

$$= 0.225 \pm (2.58)\sqrt{\frac{(0.225)(0.775)}{800}}$$

$$= 0.225 \pm 0.038$$

Probable limits $= 0.\ 187\ and\ \ 0.263$

The probable limits of the illiterate families in the population of 2000

$= 0.\ 187 * 2000\ and\ \ 0.263 * 2000$

$= 374\ and\ \ 526$

Thus, 374 to 526 are probably illiterate families.

5. In 324 throws of a six faced 'die', an odd number turned up 181 times. Is it reasonable to think that the 'die' is an unbiased one?

Sol: Probability of the turn up of an odd number , p = 3/6= ½

Since $p + q = 1$, q= ½

Expected number of successes = np

$$= 324 \times ½ = 162$$

Actual Number of heads = 181 = x

Now $x - np = 181 - 162 = 19$

Consider $z = \frac{x-\mu}{\sqrt{npq}} = \frac{x-np}{\sqrt{npq}} = \frac{19}{\sqrt{324X0.5X0.5}} = \frac{19}{9}$

z =2.11<2.58

Thus, we can say that the die is unbiased.

6. A sample of 100 days is taken from meteorological records of a certain district and 10 of them are found to be foggy. What are the probable limits of the percentage of foggy days in the district.

Sol: Proportion of foggy days in a sample of 100 days is , $p = \frac{10}{100} = 0.1$

$$\therefore\ \ q = 1 - p = 0.9 \quad , n = 100$$

Probable limits $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

$$= 0.1 \pm (2.58)\sqrt{\frac{(0.1)(0.9)}{100}}$$

$$= 0.1 \pm 0.0774$$

Probable limits $= 0.\ 0226\ and\ \ 0.1774$

Thus, the percentage of foggy days lies between 2.26% and 17.74%

7. A random sample of 500 apples was taken from a large consignment and 65 were found to be bad. Estimate the proportion of bad apples in the consignment as well as the standard error of the estimate. Also find the percentage of bad apples in the consignment.

Sol: Proportion of bad apples in the sample is given by , $p = \frac{65}{500} = 0.13$

$$\therefore \quad q = 1 - p = 0.87 \quad , n = 500$$

S.E proportion of bad apples $= \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.13)(0.87)}{500}} = 0.015$

Probable limits $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

$$= 0.13 \pm (2.58)(0.015)$$

$$= 0.13 \pm 0.0387$$

Probable limits= $0.0913 \; and \; 0.1687$

$$= 9.13\% \; and \; 16.87\%$$

Thus, the required percentage of bad apples in the consignment lies between 9.13 and 16.87.

8. The mean and standard deviation of the maximum loads supported by 60 cables are 11.09 tonnes and 0.73 tonnes respectively. Find a) 95% b) 99% confidence limits for mean of the maximum loads of all cables produced by the company.

Sol: By data $\bar{x} = 11.09$, $\sigma = 0.73$, n $= 60$

a) 95% confidence limits for mean of the maximum loads are given by

$$\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right) = 11.09 \pm (1.96)\left(\frac{0.73}{\sqrt{60}}\right)$$

$$= 11.09 \pm 0.18$$

$$= 10.91 \; and \; 11.27$$

Thus,10.91 tonnes to 11.27 tonnes are the 95% confidence limits for the mean of the maximum loads.

b) 99% confidence limits for mean of the maximum loads are given by

$$\bar{x} \pm 2.58\left(\frac{\sigma}{\sqrt{n}}\right) = 11.09 \pm (2.58)\left(\frac{0.73}{\sqrt{60}}\right)$$

$$= 11.09 \pm 0.24$$

$$= 10.85 \; and \; 11.33$$

Thus, 10.85 tonnes to 11.33 tonnes are the 95% confidence limits for the mean of the maximum loads.

9. An unbiased coin is thrown *n* times. It is desired that the relative frequency of the appearance of heads should lie between 0.49 and 0.51. Find the smallest value of *n* that will ensure this result with a) 95% confidence b) 99% confidence

Sol: Probability of getting a head, $p = \frac{1}{2}$

$$q = \frac{1}{2}$$

S.E proportion of heads $= \sqrt{\frac{pq}{n}}$

S.E proportion of heads $= \sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{n}} = \frac{1}{2\sqrt{n}}$

a) Probable limits for 95% confidence level is given by : $p \pm (1.96)\sqrt{\frac{pq}{n}}$, which should lie between 0.51 and 0.49

$$p \pm (1.96)\sqrt{\frac{pq}{n}} = 0.51 \text{ or } 0.49$$

$$0.5 \pm (1.96)\frac{1}{2\sqrt{n}} = 0.51 \text{ or } 0.49$$

$$0.5 \pm \frac{1.96}{2\sqrt{n}} = 0.51 \text{ or } 0.49$$

$0.5 + \frac{1.96}{2\sqrt{n}} = 0.51$

$\frac{1.96}{2\sqrt{n}} = 0.51 - 0.5$

$\frac{1.96}{2\sqrt{n}} = 0.01$

$0.5 - \frac{1.96}{2\sqrt{n}} = 0.49$

$\frac{1.96}{2\sqrt{n}} = 0.5 - 0.49$

$\frac{1.96}{2\sqrt{n}} = 0.01$

$\sqrt{n} = \frac{1.96}{0.02}$

$\sqrt{n} = 98$

Thus, $\mathbf{n = 9604}$

b) Taking the confidence coefficient equal to 1.645 for 90% confidence level we have as before

$$\frac{1.645}{2\sqrt{n}} = 0.01$$

$$\sqrt{n} = \frac{1.645}{0.02}$$

$$\sqrt{n} = 82.25$$

Thus, $\boldsymbol{n = 6765}$

Test of significance of a sample mean

1. A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 of them were faulty. Test his claim at a significance level of 1% and 5%.

Sol: Let p be the probability of success which being the probability of the equipment supplied to the factory conformal to the specifications.

$\therefore$ By data $p = 0.95$ and hence $q = 0.05$

$H_0: p = 0.95$ and the claim is correct.
$H_1: p < 0.95$ and the claim is false.

We choose the one tailed test to determine whether the supply is conformal to the specification.

$$\mu = np = 200 * 0.95 = 190$$

$$\sigma = \sqrt{npq} = \sqrt{200 * 0.95 * 0.05} = 3.082$$

Expected number of equipment's according to the specification =190

Actual number =182 , since 18 out of 200 were faulty.

Now, $Z = \frac{x-np}{\sigma}$

$Z = \frac{190-182}{3.082}$

$\boldsymbol{Z = 2.6}$

The value of Z is greater than the critical value 1.645 at 5% level and 2.33 at 1% level of significance.

The claim of the manufacturer (*null hypothesis that claim is correct***) is rejected at 5% as well as 1% level of significance** in accordance with the one tailed test.

2. It has been found from experience that the mean breaking strength of a particular brand of thread is 275.6gms with standard deviation of 39.7gms. Recently a sample of 36 pieces of thread showed a mean breaking strength 0f 253.2gms. Can one conclude at a significance level of a) 0.05 b) 0.01 that the thread has become inferior ?

Sol: We have to decide between the two hypothesis,

$H_0: \mu = 275.6\ gms$ **,** mean breaking strength.
$H_1: p < 275.6\ gms$, inferior in breaking strength.

We choose the one tailed test.

Mean breaking strength of a sample of 36 pieces=253.2

∴ Difference =275.6 – 253.2=22.4 ; n = 36

Now, $Z = \frac{Difference}{\left(\sigma/_{\sqrt{n}}\right)}$

$$Z = \frac{22.4}{\left(39.7/_{\sqrt{36}}\right)}$$

$Z = 3.38$

The value of Z is greater than the critical value of Z=1.645 at 5% level and 2.33 at 1% level of significance.

Under the hypothesis H_1 that the thread has become inferior is accepted at both 0.05 and 0.01 levels in accordance with one tailed test.

Test of significance of difference between means

1.In an elementary school examination the mean grade of 32 boys was 72 with a standard deviation of 8, while the mean grade of 36 girls was 75 with a standard deviation of 6. Test the hypothesis that the performance of girls is better than boys.

Sol: We have , $\bar{x}_B = 72$, $\sigma_B = 8, n_B = 32$ [Boys]

$\bar{x}_G = 75$, $\sigma_G = 6, n_G = 36$ [Girls]

Consider, $Z = \frac{\bar{x}_G - \bar{x}_B}{\sqrt{\frac{\sigma_G^2}{n_G} + \frac{\sigma_B^2}{n_B}}}$

$$Z = \frac{75-72}{\sqrt{\frac{36}{36} + \frac{64}{32}}}$$

$$Z = \sqrt{3}$$

$Z = 1.73$

$$Z = 1.73 \begin{cases} > Z_{0.05} = 1.645 \ (one\ tailed\ test) \\ < Z_{0.01} = 2.33 \ (one\ tailed\ test) \end{cases}$$

The difference in the performance of girls and boys in the examination is significant at 5% level but not at 1% level.

2. A sample of 100 bulbs produced by a company A showed a mean life of 1190 hours and a standard deviation of 90 hours. Also a sample of 75 bulbs produced by a company B showed a mean life of 1230 hours and a standard deviation of 120 hours. Is there a difference between the mean life time of the bulbs produced by the two companies at a) 5% level of significance b) 1% level of significance.

Sol: By data $\bar{x}_A = 1190$, $\sigma_A = 90, n_A = 100$ [Company A]

$$\bar{x}_B = 1230 , \sigma_B = 120, n_B = 75 \quad \text{[Company B]}$$

Consider, $Z = \frac{\bar{x}_B - \bar{x}_A}{\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}}}$

$$Z = \frac{1230 - 1190}{\sqrt{\frac{(120)^2}{75} + \frac{(90)^2}{100}}}$$

$Z = 2.42$

$$Z = 2.42 \begin{cases} > Z_{0.05} = 1.96 \; (Two\ tailed\ test) \\ < Z_{0.01} = 2.58 \; (Two\ tailed\ test) \end{cases}$$

The null hypothesis that there is no difference between the mean life time of bulbs is rejected at 5% level but not at 1% level of significance.

The null hypothesis that there is no difference between the mean life time of bulbs is rejected at both levels of significance in a one tailed test as the respective critical values are 1.645 and 2.33.

3. The mean life of two large samples of 1000 and 2000 members are 168.75 cms and 170 cms respectively. Can this be regarded as drawn from the sample population of standard deviation 6.25 cms ?

Sol: By data $\bar{x}_1 = 168.75 , n_1 = 1000$

$$\bar{x}_2 = 170 , \quad n_2 = 2000 \quad , \sigma = 6.25$$

Consider, $Z = \frac{\bar{x}_2 - \bar{x}_1}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$Z = \frac{170 - 168.75}{6.25\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$Z = 5.16$$

$Z = 5.16$ is very much greater than $Z_{0.05} = 1.96$ and $Z_{0.01} = 2.58$. Thus we say that the difference between the sample means is significant and we conclude that the samples cannot be regarded as drawn from the same population.

4. A random sample for 1000 workers in company has mean wage of Rs.50 per day and S.D of Rs.15.Another sample of 1500 workers from another company has mean wage of Rs.45 per day and S.D of Rs.20.Does the mean rate of wages varies between the two companies? Find the 95% confidence limits for the difference of the wages of the population of the two companies?

Sol: By data $\bar{x}_1 = 50 , \sigma_1 = 15, n_1 = 1000$ [Company 1]

$$\bar{x}_2 = 45 , \sigma_2 = 20, n_2 = 1500 \quad \text{[Company 2]}$$

Consider, $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$$Z = \frac{50-45}{\sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}}$$

$Z = 7.1307$

The value of $Z = 7.1307$ is greater than $Z_{0.05} = 1.96$ and $Z_{0.01} = 2.58$. Hence we can say that the difference between the mean wages is significant both at 5% and 1% level of significance.

Also 95% confidence limits for the difference of mean wages is given by

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 5 \pm 1.96(0.7012)$$

$$= 5 \pm 1.374$$

$$= 3.626 \; and \; 6.374$$

Thus we can say with 95% confidence that the difference of population mean of wages between the two companies lies between Rs.3.63 and Rs.6.37.

Student's t distribution/test

We need to test the hypothesis, whether the sample mean $\bar{x}$ differs significantly from the population mean or hypothetical value μ.

We compute, $t = \frac{\bar{x}-\mu}{s}\sqrt{n}$ and consider $|t|$.

Here , $\bar{x} = \frac{\sum x}{n}$ and $s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$v = n - 1$, denote the number of degrees of freedom 't'.

If $|t| > t_{0.05}$, the difference between $\bar{x}$ and μ is said to be significant at 5% level of significance.

If $|t| > t_{0.01}$, the difference between $\bar{x}$ and μ is said to be significant at 1% level of significance.

If $|t|$ is less than the table value at a certain level of significance, the data is said to be conformal/consistent with the hypothesis that μ is the mean of the population.

We have 95% confidence limits for μ is given by $\boldsymbol{\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.05}}$

Similarly 99% confidence limits for μ is given by $\boldsymbol{\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.01}}$

Note: 1. Confidence limits are also called **Fiducial limits**.

2. Test of significance for difference between sample means is

$$t = \frac{\bar{x}-\bar{y}}{s.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Where, $s^2 = \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{j=1}^{n_2}\left(y_j-\bar{y}\right)^2\right\}$

And degrees of freedom, $v = n_1 + n_2 - 2$

Problems

1. Find the student's t for the following variable values in a sample of eight:

-4, -2, -2, 0, 2, 2, 3, 3 taking the mean of the universe to be zero.

Sol: Wkt $t = \frac{\bar{x}-\mu}{s}\sqrt{n}$

By data $\mu = 0$ and we have $n = 8$

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

$$\bar{x} = \frac{1}{8}(-4-2-2+0+2+2+3+3)$$

$$\bar{x} = \frac{1}{4} = 0.25$$

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

$$s^2 = \frac{1}{(8-1)}\{(-4.25)^2 + (-2.25)^2 + (-2.25)^2 + (-0.25)^2 + (1.75)^2 + (1.75)^2 + (2.75)^2 + (2.75)^2\}$$

$$s^2 = \frac{1}{(7)}(49.5)$$

$$s^2 = 7.07$$

$$\therefore \boldsymbol{s = 2.66}$$

Thus, $t = \frac{0.25-0}{2.66}\sqrt{8}$

$$\boldsymbol{t = 0.266}$$

2. A machine is expected to produce nails of length 3 inches. A random sample of 25 nails gave an average length of 3.1 inch with standard deviation 0.3. Can it be said that the machine is producing nails as per specification ? ($t_{0.05}$ $for\ 24\ d.f\ is$ 2.064)

Sol: By data , we have $\mu = 3$, $\bar{x} = 3.1$, $n = 25$, $s = 0.3$

$t = \frac{\bar{x}-\mu}{s}\sqrt{n}$

$t = \frac{3.1-3}{0.3}\sqrt{25}$

$\boldsymbol{t = 1.67 < 2.064}$

Thus, the hypothesis that the machine is producing nails as per specification is accepted at 5% level of significance.

3. Ten individuals are chosen at random from a population and their heights in inches are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis that the mean height of the universe is 66 inches.($t_{0.05} = 2.262\ for\ 9\ d.f$).

Sol: We have $\mu = 66, n = 10$

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

$$\bar{x} = \frac{678}{10}$$

$$\boldsymbol{\bar{x} = 67.8}$$

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

$$s^2 = \frac{1}{9}[(63-67.8)^2 + (63-67.8)^2 + \cdots + (71-67.8)^2]$$

$$s^2 = 9.067$$

$$\boldsymbol{s = 3.011}$$

Wkt $t = \frac{\bar{x}-\mu}{s}\sqrt{n}$

$t = \frac{(67.8-66)}{3.011}\sqrt{10}$

$\boldsymbol{t = 1.89 < 2.262}$

Thus the hypothesis is accepted at 5% level of significance.

4. A sample of 10 measurements of the diameter of a sphere gave a mean of 12 cm and a standard deviation 0.15 cm. Find the 95% confidence limits for the actual diameter.

Sol: By data, $n = 10$, $\bar{x} = 12, s = 0.15$

Also $t_{0.05} = 2.262\ for\ 9\ d.f$

Confidence limits for the actual diameter is given by $\boldsymbol{\bar{x} \pm \left[\frac{s}{\sqrt{n}}\right] t_{0.05}}$

$$\bar{x} \pm \left[\frac{s}{\sqrt{n}}\right] t_{0.05} = 12 + \left[\frac{0.15}{\sqrt{10}}\right](2.262) = 12 \pm 0.1073$$

Thus 11.893 cm to 12.107 cm is the confidence limits for the actual diameter.

5. A certain stimulus administered to each of the 12 patients resulted in the following change in blood pressure 5, 2, 8, -1,3,0, 6, -2, 1, 5, 0, 4. Can it be concluded that the stimulus will increase the blood pressure ? $(t_{0.05}\ for\ 11\ d.f\ is\ 2.201)$

Sol: $\bar{x} = \frac{\sum x}{n}$

$\bar{x} = \frac{31}{12}$

$\boldsymbol{\bar{x} = 2.5833}$

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

$$s^2 = \frac{1}{11}[(5-2.58)^2 + (2-2.58)^2 + \cdots + (4-2.58)^2]$$

$s^2 = 9.538$

$\boldsymbol{s = 3.088}$

Wkt $t = \frac{\bar{x}-\mu}{s}\sqrt{n}$

Let us suppose that the stimulus administration is not accompanied with increase in blood pressure, we can take $\mu = 0$.

$t = \frac{(2.5833-0)}{3.088}\sqrt{12}$

$\boldsymbol{t = 2.8979 > 2.201}$

Hence the hypothesis is rejected at 5% level of significance. We conclude with 95% confidence that the stimulus in general is accompanied with increase in blood pressure.

6. A group of boys and girls were given an intelligence test. The mean score, S.D score and numbers in each group are as follows.

	Boys	**Girls**
Mean	74	70
S D	8	10
n	12	10

Is the difference between the means of the two groups significant at 5% level of significance ($t_{0.05}$ for 20 $d.f$ is 2.086).

Sol: By data , $\bar{x} = 74,\ s_1 = 8,\ \ n_1 = 12$ [Boys]

$\bar{y} = 70,\ s_2 = 10, n_2 = 10$ [Girls]

We have $t = \frac{\bar{x}-\bar{y}}{s.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Where, $s^2 = \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}\left(y_j - \bar{y}\right)^2\right\}$

Or $s^2 = \frac{n_1 s_1^2+n_2 s_1^2}{n_1+n_2-2}$

Now $s^2 = \frac{12(64)+10(100)}{12+10-2}$

$s^2 = \frac{1768}{20}$

$s^2 = 88.4$

$s = 9.402$

Hence $t = \frac{74-70}{(9.402)\sqrt{\frac{1}{12}+\frac{1}{10}}}$

$$\boldsymbol{t = 0.9936 < 2.086}$$

Thus the hypothesis that there is a difference between the means of the two groups is accepted at 5% level of significance.

7. A sample of 11 rats from a central population had an average blood viscosity of 3.92 with standard deviation of 0.61. On the basis of this sample, establish 95% fiducial limits for μ, the mean blood viscosity of the central population. ($t_{0.05} = 2.228$ for 10 $d.f$).

Sol: By data $\bar{x} = 3.92,\ s = 0.61, n_1 = 11$

95% fiducial limits for $\mu = \boldsymbol{\bar{x}} \pm \left[\frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}}\right] \boldsymbol{t_{0.05}}$

$\bar{x} \pm \left[\frac{s}{\sqrt{n}}\right] t_{0.05} = 3.92 \pm \left[\frac{0.61}{\sqrt{11}}\right](2.228)$

$= 3.92 \pm 0.41$

$= 3.51\ \ and\ \ 4.33$

Thus 95% confidence limits for μ are $3.51\ \ and\ \ 4.33$.

8. Two types of batteries are tested for their length of life and the following results were obtained.

Battery A : $n_1 = 10,\ \bar{x}_1 = 500\ hrs,\ s_1^2 = 100$

Battery B : $n_2 = 10,\ \bar{x}_2 = 500\ hrs,\ s_1^2 = 121$

Compute student's t and test whether there is a significant difference in the two means.

Sol: By data Battery A : $n_1 = 10,\ \bar{x}_1 = 500\ hrs,\ s_1^2 = 100$

Battery B : $n_2 = 10,\ \bar{x}_2 = 560\ hrs,\ s_1^2 = 121$

Now, $s^2 = \frac{n_1 s_1^2 + n_2 s_1^2}{n_1 + n_2 - 2}$

$$s^2 = \frac{(10*100) + (10*121)}{10+10-2}$$

$$s^2 = 122.78$$

$$\boldsymbol{s = 11.0806}$$

We have $t = \frac{\bar{x}_2 - \bar{x}_1}{s.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$t = \frac{60}{(11.0806)\sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$t = 12.1081 > 2.101 \qquad (t_{0.05} = 2.101\ for\ 18\ d.f)$$

This value of t is greater then the table value of t for 18 d.f at all levels of significance.

The null hypothesis that there is no significant difference in the two means rejected at all significance levels.

9. A group of 10 boys fed on a diet A and another group of 8 boys fed on a different diet B for a period of 6 months recorded the following increase in weight (lbs.)

Diet A : 5 6 8 1 12 4 3 9 6 10

Diet B : 2 3 6 8 10 1 2 8

Test whether diets A and B differ significantly regarding their effect on increase in weight.

Sol: Let the variable x correspond to the diet A and y to the diet B.

$$\bar{x} = \frac{\sum x}{n_1} \qquad \bar{y} = \frac{\sum y}{n_2}$$

$$\bar{x} = \frac{64}{10} \qquad \bar{y} = \frac{40}{8}$$

$$\boldsymbol{\bar{x} = 6.4} \qquad \boldsymbol{\bar{y} = 5}$$

$$\sum_{i=1}^{n_1}(x_i - \bar{x})^2 = 102.4 \quad ; \quad \sum_{j=1}^{n_2}(y_j - \bar{y})^2 = 82$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left\{\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_j - \bar{y})^2\right\}$$

$$s^2 = \frac{1}{16}\{102.4 + 82\}$$

$$s^2 = \frac{184.4}{16} = 11.525$$

$$\boldsymbol{s = 3.395}$$

We have $t = \frac{\bar{x}-\bar{y}}{s.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$t = \frac{1.4}{(3.395)\sqrt{\frac{1}{10}+\frac{1}{8}}}$$

$t = 0.869 < 2.12$ for 16 d.f

Thus we conclude that the two diets do not differ significantly regarding their effect on increase in weight.

10. Two horses A and B were tested according to the time (in seconds) to run a particular race with the following results.
Horse A : 28 30 32 33 33 29 34
Horse B : 29 30 30 24 27 29
Test whether you can discriminate between the two horses.

Sol: Let the variables x and y respectively correspond to horse A and horse B.

$$\bar{x} = \frac{\sum x}{n_1} = \frac{219}{7} = 31.3 \qquad \bar{y} = \frac{\sum y}{n_2} = \frac{169}{6} = 28.2$$

$$\sum_{i=1}^{n_1}(x_i - \bar{x})^2 = 31.43 \qquad \sum_{j=1}^{n_2}(y_j - \bar{y})^2 = 26.84$$

$$s^2 = \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_j - \bar{y})^2\right\}$$

$$s^2 = \frac{1}{11}(31.43 + 26.84)$$

$$s^2 = 5.2973$$

$$\therefore \ \boldsymbol{s = 2.3016}$$

We have $t = \frac{\bar{x}-\bar{y}}{s.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$t = \frac{31.3-28.2}{(2.3016)\sqrt{\frac{1}{7}+\frac{1}{6}}}$$

$$t = 2.42 = \begin{cases} > t_{0.05} = 2.2 \\ < t_{0.02} = 2.72 \end{cases}$$

The discrimination between the horses is significant at 5% level but not at 2% level of significance.

Chi-Square distribution

Chi-square distribution provides a measure of correspondence between the theoretical frequencies and observed frequencies.

If $O_i(i = 1,2,\dots,n)$ and $E_i(i = 1,2,\dots,n)$ respectively denotes a set of observed and estimated frequencies, the quantity **chi-square** denoted by χ^2 is defined as follows.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad ; \quad \text{degrees of freedom} = n - 1$$

Note: If the expected frequencies are less than 10, we group them suitably for computing the value of chi. Square.

Chi-square test as a test of goodness of fit

It is possible to test the hypothesis about the association of two attributes. We already discussed the fitting of Binomial distribution, Normal distribution, Poisson distribution to a given data. It is easily possible to find the theoretical frequencies from the distribution of fit.

Chi-square test helps us to test the goodness of fit of these distributions.

If the calculated value of χ^2 is less than the value of χ^2 at a specified level of significance, the hypothesis is accepted. Otherwise the hypothesis is rejected.

Problems

1.A dice is thrown 264 times and the number appearing on the face (x) follows the following frequency distribution.

x	1	2	3	4	5	6
f	40	32	28	58	54	60

Calculate the value of χ^2.

Sol: The frequencies in the given data are the observed frequencies. Assuming that the dice is unbiased, the expected number of frequencies for the numbers 1, 2, 3, 4, 5, 6 to appear on the face is $\frac{264}{6} = 44\ each$.

Now the data is as follows:

No. on the dice	1	2	3	4	5	6
Observed frequency (O_i)	40	32	28	58	54	60
Expected frequency (E_i)	44	44	44	44	44	44

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{(40-44)^2}{44} + \frac{(32-44)^2}{44} + \cdots + \frac{(60-44)^2}{44}$$

$$\chi^2 = \frac{1}{44}[16 + 144 + 256 + 196 + 100 + 256]$$

$$\chi^2 = \frac{968}{44}$$

Thus, $\boldsymbol{\chi^2 = 22}$

2. Five dice were thrown 96 times and the numbers 1, 2 or 3 appearing on the face of the dice follows the frequency distribution as below.

No. of dice showing 1,2 or 3	5	4	3	2	1	0
Frequency	7	19	35	24	8	3

Test the hypothesis that the data follows a binomial distribution. ($\chi^2_{0.05} = 11.07 \;\; for\ 5\ d.f$)

Sol: The data gives the observed frequencies and we need to calculate the expected frequencies.

Probability of a single dice throwing 1, 2 or 3 is $p = \frac{3}{6} = \frac{1}{2}$

$\therefore q = 1 - p = \frac{1}{2}$

From the binomial distribution, $P(x) = {}^{n}C_x p^x q^{(n-x)}$

The theoretical frequencies of getting 5, 4, 3, 2, 1, 0 successes with 5 dice are respectively the successive terms of the binomial expansion.

They are respectively $96 * {}^{5}C_0 * \frac{1}{2^5}$, $96 * {}^{5}C_1 * \frac{1}{2^5}$, $96 * {}^{5}C_2 * \frac{1}{2^5}$, $96 * {}^{5}C_3 * \frac{1}{2^5}$, $96 * {}^{5}C_4 * \frac{1}{2^5}$, $96 * {}^{5}C_5 * \frac{1}{2^5}$ and they are 3,15,30,30,15,3.

We have the table of observed and expected frequencies.

O_i	7	19	35	24	8	3
E_i	3	15	30	30	15	3

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{16}{3} + \frac{16}{15} + \frac{25}{30} + \frac{36}{30} + \frac{49}{15} + \frac{0}{3}$$

$$\chi^2 = 11.7 > \chi^2_{0.05} = 11.07$$

Thus the hypothesis that the data follows a binomial distribution is rejected.

3. A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured third class 90 had secured second class and 20 had secured first class. Do these figures support the general examination result which is in the ratio 4:3:2:1 for the respective categories ($\chi^2_{0.05} = 7.81$ for 3 $d.f$).

Sol: Let us take the hypothesis that these figures support to the general result in the ratio 4:3:2:1

The expected frequencies in the respective category are

$\frac{4}{10} * 500, \frac{3}{10} * 500, \frac{2}{10} * 500, \frac{1}{10} * 500$ **or** $200,\ 150,\ 100,\ 50$

We have the following table.

O_i	220	170	90	20
E_i	200	150	100	50

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{400}{200} + \frac{400}{150} + \frac{100}{100} + \frac{900}{50}$$

$$\chi^2 = 23.67 > \chi^2_{0.05} = 7.81$$

Thus the hypothesis is rejected.

4. 4 coins are tossed 100 times and the following results were obtained. Fit a binomial distribution for the data and calculate the theoretical frequencies($\chi^2_{0.05} = 9.49$ $for\ 4\ d.f$).

Number of heads	0	1	2	3	4
Frequency	5	29	36	25	5

Sol: Let x denote the number of heads and f the corresponding frequency. Since the data is in the form of a frequency distribution we shall first calculate the mean.

Mean, $\mu = \frac{\sum fx}{\sum f} = \frac{(0)(5)+(1)(29)+(2)(36)+(3)(25)+(4)(5)}{5+29+36+25+5} = \frac{196}{100} = 1.96$

From binomial distribution, $\mu = np$. Here n=4

Hence $4p = 1.96$

$$\boldsymbol{p = 0.49}$$

$$\boldsymbol{q = 0.51}$$

From binomial distribution,
$P(x) = nc_x p^x q^{n-x}$

$\boldsymbol{P(x) = 4c_x (0.49)^x (0.51)^{4-x}}$
Since 4 coins were tossed , expected(theoretical) frequencies are obtained from
$\mathrm{F}(x) = 100P(x)$

$\mathrm{F}(x) = 100.4c_x(0.49)^x(0.51)^{4-x}$ where $x = 0,1,2,3,4.$

$\mathrm{F}(0) = 100.4c_0(0.49)^0(0.51)^4 = 6.765 \cong 7$

$\mathrm{F}(1) = 100.4c_1(0.49)^1(0.51)^3 = 25.999 \cong 26$

$\mathrm{F}(2) = 100.4c_2(0.49)^2(0.51)^2 = 37.47 \cong 37$

$\mathrm{F}(3) = 100.4c_3(0.49)^3(0.51)^1 = 24.0004 \cong 24$

$\mathrm{F}(4) = 100.4c_4(0.49)^4(0.51)^0 = 5.765 \cong 6$

Thus the required theoretical frequencies are **7,26,37,24,6**.

We have the following table

O_i	5	29	36	25	5
E_i	7	26	37	24	6

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$\chi^2 = \frac{4}{7} + \frac{9}{26} + \frac{1}{37} + \frac{1}{24} + \frac{1}{6}$

$\chi^2 = 1.15 < \chi^2_{0.05} = 9.49$

Thus the hypothesis that the fitness is good can be accepted.

5. Fit a poisson distribution for the following data and test the goodness of fit given that $\chi^2_{0.05} = 7.815 \ for\ 3\ d.f$

x	**0**	**1**	**2**	**3**	**4**
f	122	60	15	2	1

Sol: Let x denote the number of heads and f the corresponding frequency. Since the data is in the form of a frequency distribution we shall first calculate the mean.

Mean, $\mu = \frac{\sum fx}{\sum f} = \frac{(0)(122)+(1)(60)+(2)(15)+(3)(2)+(4)(1)}{122+60+15+2+1} = \frac{100}{200} = 0.2$

From poisson distribution, $\mu = m$.

$$\therefore\ \boldsymbol{m = 0.2}$$

Wkt, $P(x) = \frac{m^x e^{-m}}{x!}$

Let $F(x) = 200 * P(x)$

$F(x) = 200 * \frac{m^x e^{-m}}{x!}$

$F(x) = 200 * \frac{(0.5)^x e^{-(0.5)}}{x!}$ But $e^{-(0.5)} = 0.6065$

$F(x) = 121.3 \frac{(0.5)^x}{x!}$

$F(0) = 121$, $F(1) = 61$, $F(2) = 15$, $F(3) = 3$, $F(4) = 0$

Thus the required theoretical frequencies are 121,61,15,3,0.

Since the last of the expected frequency is 0 we shall club it with the previous one.

We have the following table

O_i	122	60	15	2+1=3
E_i	121	61	15	3+0=3

$$\boldsymbol{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

$\chi^2 = \frac{1}{121} + \frac{1}{61} + 0 + 0$

$\chi^2 = 0.025 < \chi^2_{0.05} = 7.815$

Thus the hypothesis that the fitness is good can be accepted.

6. The number of accidents per day (x) as recorded in a textile industry over a period of 400 days is given below. Test the goodness of fit in respect of Poisson distribution of fit to the given data$(\chi^2_{0.05} = 9.49 \ \ for\ 4\ d.f)$.

x	0	1	2	3	4	5
f	173	168	37	18	3	1

Sol: Let x denote the number of heads and f the corresponding frequency. Since the data is in the form of a frequency distribution we shall first calculate the mean.

Mean, $\mu = \frac{\sum fx}{\sum f} = \frac{(0)(173)+(1)(168)+(2)(37)+(3)(18)+(4)(3)+(5)(1)}{173+168+37+18+3+1} = \frac{313}{400} = 0.7825$

From poisson distribution, $\mu = m$.

$$\therefore\ \boldsymbol{m = 0.7825}$$

Wkt , $P(x) = \frac{m^x e^{-m}}{x!}$

Let $\ F(x) = 400 * P(x)$

$F(x) = 400 * \frac{m^x e^{-m}}{x!}$

$F(x) = 400 * \frac{(0.7825)^x e^{-(0.7825)}}{x!}$ But $e^{-(0.7825)} = 0.4573$

$F(x) = 182.92\, \frac{(0.7825)^x}{x!}$

$F(0) = 183$, $F(1) = 143$, $F(2) = 56$, $F(3) = 15$, $F(4) = 3$, $F(5) = 0$
Thus the required theoretical frequencies are 183,143,56,15,3,0.

Since the last of the expected frequency is 0 we shall club it with the previous one.

We have the following table

$\boldsymbol{O_i}$	173	168	37	18	3+1=4
E_i	183	143	56	15	3+0=3

$$\boldsymbol{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

$$\chi^2 = \frac{100}{183} + \frac{625}{143} + \frac{361}{56} + \frac{9}{15} + \frac{1}{3}$$

$\chi^2 = 12.297 > \chi^2_{0.05} = 9.49$

Thus the hypothesis that the fitness is good can be rejected.

MODULE-5
DESIGN OF EXPERIMENTS &ANOVA

ANALYSIS OF VARIANCE

The analysis of variance is a powerful statistical tool for tests of significance.

We consider the following types of ANOVA

I) One- way classification (One factor ANOVA) –Completely Randomized Design.
II) Two-Way classification (Two factors)-Randomized Block Design.
III) Latin Square Design-Three Way ANOVA

ONE - WAY ANOVA

One- way classification observations are classified according to one factor

WORKING PROCEDURE:

Null Hypothesis H_0: $There\ is\ no\ significance\ difference\ between\ \dots\dots\dots$

Alternative Hypothesis H_1: $There\ is\ significance\ difference\ between\ \dots\dots\dots..$

STEP1: To find the following

$$N = Number\ of\ observations$$

$$T = \sum X_1 + \sum X_2 + \sum X_3 + \cdots$$

Correction factor: $C.F = \frac{T^2}{N}$

STEP2: To Calculate the following

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \cdots \quad - C.F$$

$$SSC = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots \; - C.F$$

$$SSE = TSS - SSC$$

STEP3: O NE- WAY ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	VARIATION RATIO
BETWEEN COLUMNS	SSC =	C – 1 =	$MSC = \frac{SSC}{C-1}$	$F_c = \frac{MSE}{MSC}$
ERROR	SSE =	N – C =	$MSE = \frac{SSE}{N-C}$	
TOTAL				

NOTE: IF CALCULATED VALUE OF F < TABULATED VALUE OF F, THEN H_0 IS **ACCEPTED**.

IF CALCULATED VALUE OF F > TABULAED VALUE OF F, THEN H_0 $IS\ REJECTED$.

PROBLEM 1:

The following figures relate to production in kgs. of three variables A, B, C of wheat sown on 12 plots

A	14	16	18		
B	14	13	15	22	
C	18	16	19	19	22

Is there any significant difference in the production of the Varieties.

SOLUTION:

Null Hypothesis:

H_0 : $There\ is\ no\ significance\ difference\ between\ in\ the\ production\ of\ varieties.$

Alternative Hypothesis:

$\boldsymbol{H_1}$: $There\ is\ significance\ difference\ between\ in\ the\ production\ of\ the\ varieties.$

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2
14	14	18	196	196	324
16	13	1	256	169	256
18	15	19	324	225	341
	22	19		484	361
		20			400
TOTAL $\sum X_1 = 48$	$\sum X_2 = 64$	$\sum X_3 = 92$	$\sum X_1^2 = 776$	$\sum X_2^2 = 1074$	$\sum X_3^2 = 1702$

STEP1: To find the following

$$N = Number\ of\ observations = 12$$

$$T = \sum X_1 + \sum X_2 + \sum X_3 = 48 + 64 + 92 = 204$$

Correction factor: $C.F = \frac{T^2}{N} = \frac{(204)^2}{12} = 3468$

STEP2: To Calculate the following

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 - C.F$$

$$= 776 + 1074 + 1702 - 3468 = 84$$

$$SSC = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} - C.F$$

$$= \frac{(48)^2}{3} + \frac{(64)^2}{4} + \frac{(92)^2}{5} - 3468 = 16.8$$

$$SSE = TSS - SSC = 84 - 16.8 = 67.2$$

STEP 3: ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	VARIATION RATIO
BETWEEN COLUMNS	SSC = 16.8	C – 1 = 3- 1 = 2	$MSC = \frac{SSC}{C-1}$ $= \frac{16.8}{2}$ $= \mathbf{8.4}$	$F_c = \frac{MSC}{MSE}$ $= \frac{8.4}{7.47}$ $= 1.13$
ERROR	SSE = 67.2	N – C = 12 – 3 = 9	$MSE = \frac{SSE}{N-C}$ $= \frac{67.2}{9} = 7.47$	
TOTAL				

RESULT:

Table value of F_c at 5% level in (2 , 9) degrees of freedom (d.f) is **4.26**

Since calculated value of F_c $< tabulated\ value\ of\ F_c$

Which implies H_0 $is\ accepted.$

TWO WAY ANOVA

Two - way classification observations are classified according to two factors

WORKING PROCEDURE:

Null Hypothesis H_0: $There\ is\ no\ significance\ difference\ between\ \dots\dots\dots$

Alternative Hypothesis H_1: $There\ is\ significance\ difference\ between\ \dots\dots\dots..$

STEP1: To find the following

$$N = Number\ of\ observations$$

$$T = \sum X_1 + \sum X_2 + \sum X_3 + \cdots$$

Correction factor: $C.F = \frac{T^2}{N}$

STEP2: To Calculate the following

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \cdots \quad - C.F$$

$$SSC = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots \quad - C.F$$

$$SSC = \frac{(\sum Y_1)^2}{n_1} + \frac{(\sum Y_2)^2}{n_2} + \cdots \quad - C.F$$

$$SSE = TSS - SSC - SSR$$

STEP 3: TWO – WAY ANOVA TABLE:

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	VARIATION RATIO
BETWEEN COLUMNS	SSC =	C – 1 =	$MSC = \frac{SSC}{C-1}$	$F_c = \frac{MSE}{MSC}$
BETWEEN ROWS	SSR =	R – 1 =	$MSR = \frac{SSR}{R-1}$	$F_r = \frac{MSE}{MSR}$
ERROR	SSE =	(C – 1) X (R – 1) =	$MSE = \frac{SSE}{N-C}$	

PROBLEM 2:

Four doctors each test four treatments for a certain disease and observe the number of days each patient takes to recover. The results are as follows (recovery time in days)

	Treatment			
Doctor	T1	T2	T3	T4
D1	10	14	19	20
D2	11	15	17	21
D3	9	12	16	19
D4	8	13	17	20

Discuss the difference between (a) doctors and (b) treatments.

SOLUTION:

Null Hypothesis: $\boldsymbol{H}_0$: $There\ is\ no\ significance\ difference\ between\ in\ the\ perfomence\ of\ doctors\ and\ treatments.$

Alternative Hypothesis:

$\boldsymbol{H}_1$: $There\ is\ significance\ difference\ between\ in\ the\ in\ the\ perfomence\ of\ doctors\ and\ treatments$

USE CODE: 15

X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
-5	-1	4	5	$\sum Y_1 = 3$	25	1	16	25
-4	0	2	6	$\sum Y_2 = 4$	16	0	4	36
-6	-3	1	4	$\sum Y_3 = -4$	36	9	1	16
-7	-2	2	5	$\sum Y_4 = -2$	49	4	4	25
TOTAL $\sum X_1 = -22$	$\sum X_2 = -6$	$\sum X_3 = 9$	$\sum X_4 = 20$		$\sum X_1^2 = 126$	$\sum X_2^2 = 14$	$\sum X_3^2 = 25$	$\sum X_4^2 = 102$

STEP1: To find the following

$$N = Number\ of\ observations = 16$$

$$T = \sum X_1 + \sum X_2 + \sum X_3 + \sum X_4$$

$$= -22 - 6 + 9 + 20 = 1$$

Correction factor: $C.F = \frac{T^2}{N} = \frac{(1)^2}{16} = \mathbf{0.0625}$

STEP2: To Calculate the following

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F$$

$$= 126 + 14 + 25 + 102 - 0.0625$$

$$= 266.94$$

$$SSC = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} + \frac{(\sum X_4)^2}{n_4} - C.F$$

$$= \frac{(-22)^2}{4} + \frac{(-6)^2}{4} + \frac{(9)^2}{4} + \frac{(20)^2}{4} - 0.0625$$

$$= 250.19$$

$$SSR = \frac{(\sum Y_1)^2}{n_1} + \frac{(\sum Y_2)^2}{n_2} + \frac{(\sum Y_3)^2}{n_3} + \frac{(\sum Y_4)^2}{n_4} - C.F$$

$$= \frac{(3)^2}{4} + \frac{(4)^2}{4} + \frac{(-4)^2}{4} + \frac{(-2)^2}{4} - 0.0625$$

$$= 11.19$$

$$\boldsymbol{SSE = TSS - SSC - SSR}$$

$$= 266.94 - 250.19 - 11.19$$

$$= 5.56$$

STEP 3: ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	VARIATION RATIO
BETWEEN COLUMNS	SSC = 250.19	C – 1 = 4 - 1 = 3	$MSC = \frac{SSC}{C-1}$ $= \frac{250.19}{3}$ = 83.40	$F_c = \frac{MSC}{MSR}$ $= \frac{83.40}{0.62}$ = **134.52**
BETWEEN ROWS	SSR = 11.19	R – 1 = 4 – 1 = 3	$MSR = \frac{SSR}{R-1}$ $= \frac{11.19}{3}$ = 3.73	$F_r = \frac{MSR}{MSE}$ $= \frac{3.73}{0.62}$ $\mathbf{= 6.02}$
ERROR	SSE = 5.56	(C – 1) X(R – 1) = 9	$MSE = \frac{SSE}{N-C}$ $= \frac{5.56}{9}$ = 0.62	

TOTAL

From the F- table, the value of F at 5% level in (3, 9) d.f is **3.86**

RESULT:

Since both calculated values of F_c *and* $F_r >$ *table value of* F

Which implies H_0 *is Rejected*.

Hence the difference between doctors is significant and that between the treatments is highly significant.

LATIN SQUARE DESIGN:

Three -way classification observations are classified according to three factors

WORKING PROCEDURE:

Null Hypothesis $H_0: There\ is\ no\ significance\ difference\ between \dots \dots \dots$

Alternative Hypothesis $H_1: There\ is\ significance\ difference\ between \dots \dots \dots ..$

STEP1: To find the following

$$N = Number\ of\ observations$$

$$T = \sum X_1 + \sum X_2 + \sum X_3 + \cdots$$

Correction factor: $C.F = \frac{T^2}{N}$

STEP2: To Calculate the following

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \cdots \quad - C.F$$

$$SSC = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \cdots \quad - C.F$$

$$SSR = \frac{(\sum Y_1)^2}{n_1} + \frac{(\sum Y_2)^2}{n_2} + \cdots \quad - C.F$$

$$SST = \frac{(\sum Z_1)^2}{n_1} + \frac{(\sum Z_2)^2}{n_2} + \cdots \quad - C.F$$

$$SSE = TSS - SSC - SSR - SST$$

STEP 3: ANOVA TABLE – LATIN SQUARE DESIGN

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	VARIATION RATIO
BETWEEN COLUMNS	SSC =	K – 1 =	$MSC = \frac{SSC}{C-1}$	$F_c = \frac{MSE}{MSC}$
BETWEEN ROWS	SSR =	K – 1 =	$MSR = \frac{SSR}{R-1}$	$F_r = \frac{MSE}{MSR}$
BETWEEN TREATMENTS	SST =	K – 1 =	$MST = \frac{SST}{N-C}$	$F_T = \frac{MSE}{MST}$
ERROR	SSE =	(K – 1) X (K – 2) =	$MSE = \frac{SSE}{(\mathrm{K}-1)\mathrm{X}\,(\mathrm{K}-2)}$	

TOTAL

PROBLEM 3:

Analysis the variance in the following **Latin square** of yields (in kgs) of wheat where A, B,C,D denote different methods of cultivation

D122	A121	C123	B122
B124	C123	A122	D125
A120	B119	D120	C121
C122	D123	B121	A122

Examine whether the different methods of cultivation have given significantly different yields.

SOLUTION:

Null Hypothesis: H_0 :
$There\ is\ no\ significance\ difference\ between\ in\ the\ Letters\ and\ the\ different\ method\ of\ cultivation.$

Alternative Hypothesis:

$H_1 : There\ is\ significance\ difference\ between\ in\ the\ in\ the\ \ Letters\ and\ the\ different\ method\ of\ cultivation.$

USE CODE: 120

X_1	X_2	X_3	X_4	TOTAL	X_1^2	X_2^2	X_3^2	X_4^2
2	1	3	2	$\sum Y_1 = 8$	4	1	9	4
4	3	2	5	$\sum Y_2 = 14$	16	4	4	25
0	- 1	0	1	$\sum Y_3 = 0$	0	1	0	1
2	3	1	2	$\sum Y_4 = 8$	4	9	1	4
TOTAL $\sum X_1 = 8$	$\sum X_2 = 6$	$\sum X_3 = 6$	$\sum X_4 = 10$		$\sum X_1^2 = 24$	$\sum X_2^2 = 15$	$\sum X_3^2 = 14$	$\sum X_4^2 = 34$

A	B	C	D
1	**2**	**3**	**2**
2	**4**	**3**	**5**
0	**-1**	**1**	**0**
2	**1**	**2**	**3**

$\sum Z1 = 5$ $\sum Z2=6$ $\sum Z3=9$ $\sum Z4=10$

STEP1: To find the following

$$N = Number\ of\ observations = 16$$

$$T = \sum X_1 + \sum X_2 + \sum X_3 + \sum X_4 = 8 + 6 + 6 + 10 = 30$$

Correction factor: $C.F = \frac{T^2}{N} = \frac{(30)^2}{16} = 56.25$

STEP2: To Calculate the following

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F$$

$$= 24 + 15 + 14 + 34 - 56.25$$

$$= 35.75$$

$$SSC = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} + \frac{(\sum X_4)^2}{n_4} - C.F$$

$$= \frac{(8)^2}{4} + \frac{(6)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 56.25$$

$$= 2.75$$

$$SSR = \frac{(\sum Y_1)^2}{n_1} + \frac{(\sum Y_2)^2}{n_2} + \frac{(\sum Y_3)^2}{n_3} + \frac{(\sum Y_4)^2}{n_4} - C.F$$

$$= \frac{(8)^2}{4} + \frac{(14)^2}{4} + \frac{(0)^2}{4} + \frac{(8)^2}{4} - 56.25$$

$$= 24.75$$

$$SST = \frac{(\sum Z_1)^2}{n_1} + \frac{(\sum Z_2)^2}{n_2} + \frac{(\sum Z_3)^2}{n_3} + \frac{(\sum Z_4)^2}{n_4} - C.F$$

$$= \frac{(5)^2}{4} + \frac{(6)^2}{4} + \frac{(9)^2}{4} + \frac{(10)^2}{4} - 56.25$$

$$= 4.25$$

$$SSE = TSS - SSC - SSR - SST$$

$$= 35.75 - 2.75 - 24.75 - 4.25$$

$$= 4.0$$

STEP 3: ANOVA TABLE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES	VARIATION RATIO
BETWEEN COLUMNS	SSC = 2.75	K – 1 = 4 - 1 = 3	$MSC = \frac{SSC}{K-1}$ $= \frac{2.75}{3}$ = 0.92	$F_c = \frac{MSC}{MSR}$ $= \frac{0.92}{0.67}$ = 1.37313
BETWEEN ROWS	SSR = 24.75	K – 1 = 4 – 1 = 3	$MSR = \frac{SSR}{K-1}$ $= \frac{24.75}{3}$ = 8.25	$F_r = \frac{MSR}{MSE}$ $= \frac{8.25}{0.67}$ $= \mathbf{6.02}$
BETWEEN TREATMENTS	SST = 4.25	K – 1 = 4 – 1 = 3	$MST = \frac{SST}{K-1}$ $= \frac{4.25}{3}$ = 1.42	$F_T = \frac{MST}{MSE}$ $= \frac{1.42}{0.67}$ $= \mathbf{2.1194}$
ERROR	SSE = 4.0	(K – 1) X(K – 2) = 6	$MSE = \frac{SSE}{D.F}$ $= \frac{4.0}{6}$ = 0.67	
TOTAL				

From the F- table, the value of F at 5% level in (3, 6) d.f is **4.76**

RESULT:

Since both calculated values of F_c and $F_T < table\ value\ of\ F$

Since with respect to letters, the difference between the method of cultivation is not significant.

.

THANK YOU