SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

SCHOOL OF SCIENCE AND HUMANITIES

DEPARTMENT OF MATHEMATICS

# UNIT – I - Probability Concepts and Random Variable – SMTA1402

# Sathyabama Institute of Science and Technology

**SMTA1402 - Probability and Statistics**

**Unit-1 Probability Concepts and Random Variable**

## Random Experiment

An experiment whose outcome or result can be predicted with certainty is called a Deterministic experiment.

Although all possible outcomes of an experiment may be known in advance the outcome of a particular performance of the experiment cannot be predicted owing to a number of unknown causes. Such an experiment is called a Random experiment.

(e.g.) Whenever a fair dice is thrown, it is known that any of the 6 possible outcomes will occur, but it cannot be predicted what exactly the outcome will be.

## Sample Space

The set of all possible outcomes which are assumed equally likely.

## Event

A sub-set of S consisting of possible outcomes.

## Mathematical definition of Probability

Let S be the sample space and A be an event associated with a random experiment. Let $n(S)$ and $n(A)$ be the number of elements of S and A. then the probability of event A occurring is denoted as P(A), is denoted by

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

Note: 1. It is obvious that $0 \le P(A) \le 1$.
2. If A is an impossible event, $P(A) = 0$.
3. If A is a certain event , $P(A) = 1$.

A set of events is said to be mutually exclusive if the occurrence of any one them excludes the occurrence of the others. That is, set of the events does not occur simultaneously,

$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n, \ldots) = 0$ A set of events is said to be mutually exclusive if the occurrence of any one them excludes the occurrence of the others. That is, set of the events does not occur simultaneously,

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n, \ldots) = 0$$

## Axiomatic definition of Probability

Let S be the sample space and A be an event associated with a random experiment. Then the probability of the event A, P(A) is defined as a real number satisfying the following axioms.

1. $0 \le P(A) \le 1$
2. $P(S) = 1$
3. If A and B are mutually exclusive events, $P(A \cup B) = P(A) + P(B)$ and
4. If $A_1, A_2\, A_3, \ldots, A_n, \ldots$ are mutually exclusive events,

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n, \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) \ldots$$

**Important Theorems**

**Theorem 1:** Probability of impossible event is zero.

**Proof:** Let S be sample space (certain events) and $\phi$ be the impossible event.

Certain events and impossible events are mutually exclusive.

$P(S \cup \phi) = P(S) + P(\phi)$ (Axiom 3)

$S \cup \phi = S$

$P(S) = P(S) + P(\phi)$

$P(\phi) = 0$, hence the result.

**Theorem 2:** If $\overline{A}$ is the complementary event of $A$, $P(\overline{A}) = 1 - P(A) \le 1$.

**Proof:** Let $A$ be the occurrence of the event

$\overline{A}$ be the non-occurrence of the event .

Occurrence and non-occurrence of the event are mutually exclusive.

$P(A \cup \overline{A}) = P(A) + P(\overline{A})$

$A \cup \overline{A} = S \quad \Rightarrow \quad P(A \cup \overline{A}) = P(S) = 1$

$\therefore \ 1 = P(A) + P(\overline{A})$

$P(\overline{A}) = 1 - P(A) \le 1$.

**Theorem 3:** (Addition theorem)

If A and B are any 2 events,

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

**Proof:** We know, $A = A\overline{B} \cup AB$ and $B = \overline{A}B \cup AB$

$\therefore \quad P(A) = P(A\overline{B}) + P(AB)$ and $P(B) = P(\overline{A}B) + P(AB)$ (Axiom 3)

$P(A) + P(B) = P(A\overline{B}) + P(AB) + P(\overline{A}B) + P(AB)$

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

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

Note: The theorem can be extended to any 3 events, A,B and C

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

**Theorem 4:** If $B \subset A$, $P(B) \le P(A)$.

**Proof:** $A$ and $A\overline{B}$ are mutually exclusive events such that $B \cup A\overline{B} = A$

$\therefore \quad P(B \cup A\overline{B}) = P(A)$

$P(B) + P(A\overline{B}) = P(A)$ (Axiom 3)

$P(B) \le P(A)$

**Conditional Probability**

The conditional probability of an event B, assuming that the event A has happened, is denoted by P(B/A) and defined as

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, \text{ provided } P(A) \ne 0$$

**Product theorem of probability**

Rewriting the definition of conditional probability, We get

$$P(A \cap B) = P(A)P(A/B)$$

The product theorem can be extended to 3 events, A, B and C as follows:

$$P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B)$$

Note: 1. If $A \subset B$, $P(B/A) = 1$, since $A \cap B = A$.

2. If $B \subset A$, $P(B/A) \geq P(B)$, since $A \cap B = B$, and $\frac{P(B)}{P(A)} \geq P(B)$,

As $P(A) \leq P(S) = 1$.

3. If A and B are mutually exclusive events, $P(B/A) = 0$, since $P(A \cap B) = 0$.
4. If $P(A) > P(B)$, $P(A/B) > P(B/A)$.
5. If $A_1 \subset A_2$, $P(A_1/B) \leq P(A_2/B)$.

**Independent Events**

A set of events is said to be independent if the occurrence of any one of them does not depend on the occurrence or non-occurrence of the others.

If the two events A and B are independent, the product theorem takes the form $P(A \cap B) = P(A) \times P(B)$, Conversely, if $P(A \cap B) = P(A) \times P(B)$, the events are said to be independent (pair wise independent).

The product theorem can be extended to any number of independent events, If $A_1\ A_2\ A_3 \ldots A_n$ are $n$ independent events, then

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = P(A_1) \times P(A_2) \times P(A_3) \times \ldots \times P(A_n)$$

**Theorem 4:**

If the events A and B are independent, the events $\bar{A}$ and B are also independent.

**Proof:**

The events $A \cap B$ and $\bar{A} \cap B$ are mutually exclusive such that $(A \cap B) \cup (\bar{A} \cap B) = B$

$\therefore \quad P(A \cap B) + P(\bar{A} \cap B) = P(B)$

$P(\bar{A} \cap B) = P(B) - P(A \cap B)$

$= P(B) - P(A)\,P(B)$ ($\therefore$ A and B are independent)

$= P(B)\,[1 - P(A)]$

$= P(\bar{A})\,P(B)$.

**Theorem 5:**

If the events A and B are independent, the events $\bar{A}$ and $\bar{B}$ are also independent.

**Proof:**

$P(\bar{A} \cap \bar{B}) = P\left(\overline{A \cup B}\right) = 1 - P(A \cup B)$

$= 1 - [\,P(A) + P(B) - P(A \cap B)]$ (Addition theorem)

$= [1 - P(A)] - P(B)\,[1 - P(A)]$

$= P(\bar{A})P(\bar{B})$.

**Problem 1:**

From a bag containing 3 red and 2 balck balls, 2 ball are drawn at random. Find the probability that they are of the same colour.

**Solution :**

Let A be the event of drawing 2 red balls

B be the event of drawing 2 black balls.

$\therefore \quad P(A \cup B) = P(A) + P(B)$

$$= \frac{3C_2}{5C_2} + \frac{2C_2}{5C_2} = \frac{3}{10} + \frac{1}{10} = \frac{2}{5}$$

**Problem 2:**

When 2 card are drawn from a well-shuffled pack of playing cards, what is the probability that they are of the same suit?

**Solution :**

Let A be the event of drawing 2 spade cards

B be the event of drawing 2 claver cards

C be the event of drawing 2 hearts cards

D be the event of drawing 2 diamond cards.

$$\therefore P(A \cup B \cup C \cup D) = 4\frac{13C_2}{52C_2} = \frac{4}{17}.$$

**Problem 3:**

When A and B are mutually exclusive events such that P(A) = 1/2 and P(B) = 1/3, find $P(A \cup B)$ and $P(A \cap B)$.

**Solution :**

$P(A \cup B) = P(A) + P(B) = 5/6$ ; $P(A \cap B) = 0$.

**Problem 4:**

If P(A) = 0.29, P(B) = 0.43, find $P(A \cap \overline{B})$, if A and B are mutually exclusive.

**Solution :**

We know $A \cap \overline{B} = A$

$$P(A \cap \overline{B}) = P(A) = 0.29$$

**Problem 5:**

A card is drawn from a well-shuffled pack of playing cards. What is the probability that it is either a spade or an ace?

**Solution :**

Let A be the event of drawing a spade

B be the event of drawing a ace

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

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{4}{13}.$$

**Problem 6:**

If P(A) = 0.4, P(B) = 0.7 and P(A ∩ B) = 0.3, find $P(\bar{A} \cap \bar{B})$.

**Solution :**

$$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$$
$$= 1 - [P(A) + P(B) - P(A \cap B)]$$
$$= 0.2$$

**Problem 7:**

If P(A) = 0.35, P(B) = 0.75 and P(A ∪ B) = 0.95, find $P(\bar{A} \cup \bar{B})$.

**Solution :**

$$P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B) = 1 - [P(A) + P(B) - P(A \cup B)] = 0.85$$

**Problem 8:**

A lot consists of 10 good articles, 4 with minor defects and 2 with major defects. Two articles are chosen from the lot at random(with out replacement). Find the probability that (i) both are good, (ii) both have major defects, (iii) at least 1 is good, (iv) at most 1 is good, (v) exactly 1 is good, (vi) neither has major defects and (vii) neither is good.

**Solution :**

(i) P( both are good) = $\frac{10C_2}{16C_2} = \frac{3}{8}$

(ii) P(both have major defects) = $\frac{2C_2}{16C_2} = \frac{1}{120}$

(iii) P(at least 1 is good) = $\frac{10C_1 6C_1 + 10C_2}{16C_2} = \frac{7}{8}$

(iv) P(at most 1 is good) = $\frac{10C_0 6C_2 + 10C_1 6C_1}{16C_2} = \frac{5}{8}$

(v) P(exactly 1 is good) = $\frac{10C_1 6C_1}{16C_2} = \frac{1}{2}$

(vi) P(neither has major defects) = $\frac{14C_2}{16C_2} = \frac{91}{120}$

(vii) P(neither is good) = $\frac{6C_2}{16C_2} = \frac{1}{8}$.

**Problem 9:**

If A, B and C are any 3 events such that P(A) = P(B) = P(C) = 1/4, P(A ∩ B) = P(B ∩ C) = 0; P(C ∩ A) = 1/8. Find the probability that at least 1 of the events A, B and C occurs.

**Solution :**

Since P(A ∩ B) = P(B ∩ C) = 0; P(A ∩ B ∩ C) = 0

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$$
$$= \frac{3}{4} - 0 - 0 - \frac{1}{8} = \frac{5}{8}.$$

**Problem 10:**

A box contains 4 bad and 6 good tubes. Two are drawn out from the box at a time. One of them is tested and found to be good. What is the probability that the other one is also good?

**Solution :**

Let A be a good tube drawn and B be an other good tube drawn.

P(both tubes drawn are good) = P(A $\cap$ B) = $\frac{6C_2}{10C_2} = \frac{1}{3}$

P(B/A) = $\frac{P(A \cap B)}{P(A)} = \frac{1/3}{6/10} = \frac{5}{9}$ (By conditional probability)

**Problem 11:**

In shooting test, the probability of hitting the target is 1/2, for a, 2/3 for B and ¾ for C. If all of them fire at the target, find the probability that (i) none of them hits the target and (ii) at least one of them hits the target.

**Solution :**

Let A, B and C be the event of hitting the target .
P(A) = 1/2, P(B) = 2/3, P(C) = 3/4
P($\overline{A}$) = 1/2, P($\overline{B}$) = 1/3, P($\overline{C}$) = 1/4

P(none of them hits) = P($\overline{A} \cap \overline{B} \cap \overline{C}$) = P($\overline{A}$) × P($\overline{B}$) × P($\overline{C}$) = 1/24

P(at least one hits) = 1 – P(none of them hits)
= 1 – (1/24) = 23/24.

**Problem 12:**

A and B alternatively throw a pair of dice. A wins if he throws 6 before B throws 7 and B wins if he throws 7 before A throws 6. If A begins, show that his chance of winning is 30/61.

**Solution :**

Let A be the event of throwing 6
B be the event of throwing 7.

P(throwing 6 with 2 dice) = 5/36
P(not throwing 6) = 31/36

P(throwing 7 with 2 dice) = 1/6
P(not throwing 7) = 5/6

A plays in I, III, V,……trials.
A wins if he throws 6 before Be throws 7.

P(A wins) = P(A $\cup$ $\overline{A}\,\overline{B}$A $\cup$ $\overline{A}\,\overline{B}\,\overline{A}\,\overline{B}$A $\cup$ …… )
= P(A) + P($\overline{A}\,\overline{B}$A) + P($\overline{A}\,\overline{B}\,\overline{A}\,\overline{B}$A) + ……

$$= \frac{5}{36} + \left(\frac{31}{36} \times \frac{5}{6}\right)\frac{5}{36} + \left(\frac{31}{36} \times \frac{5}{6}\right)^2 \frac{5}{36} + \dots$$

$$= \frac{30}{61}$$

**Problem 13:**

A and B toss a fair coin alternatively with the understanding that the first who obtain the head wins. If A starts, what is his chance of winning?

**Solution :**

P(getting head) = 1/2 ,  P(not getting head) = 1/2

A plays in I, III, V,……trials.
A wins if he gets head before B.

$$P(\text{A wins}) = P(A \cup \bar{A}\,\bar{B}\,A \cup \bar{A}\,\bar{B}\,\bar{A}\,\bar{B}\,A \cup \ldots\ldots)$$
$$= P(A) + P(\bar{A}\,\bar{B}\,A) + P(\bar{A}\,\bar{B}\,\bar{A}\,\bar{B}\,A) + \ldots\ldots$$
$$= \frac{1}{2} + \left(\frac{1}{2}\times\frac{1}{2}\right)\frac{1}{2} + \left(\frac{1}{2}\times\frac{1}{2}\right)^2\frac{1}{2} + \cdots$$
$$= \frac{2}{3}$$

**Problem 14:**

A problem is given to 3 students whose chances of solving it are 1/2 , 1/3 and 1/4 . What is the probability that (i) only one of them solves the problem and (ii) the problem is solved.

**Solution :**

P( A solves) = 1/2 P(B) = 1/3  P(C) = 1/4
$P(\bar{A}) = 1/2$, $P(\bar{B}) = 2/3$, $P(\bar{C}) = 3/4$

P(none of them solves) = $P(\bar{A} \cap \bar{B} \cap \bar{C}) = P(\bar{A}) \times P(\bar{B}) \times P(\bar{C}) =$ 1/4

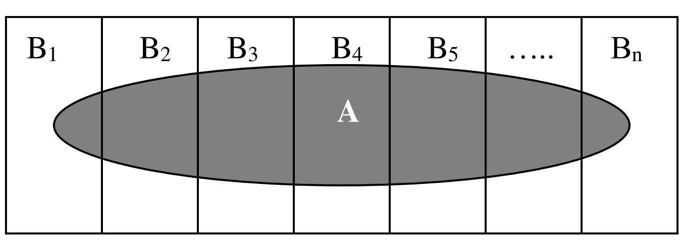
P(at least one solves) = 1 – P(none of them solves)
= 1 – (1/4) = 3/4 .

**Baye's Theorem**

Statement: If $B_1, B_2, B_3, \ldots B_n$ be a set of exhaustive and mutually exclusive events associated with a random experiment and A is another event associated with $B_i$, then

$$P(B_i / A) = \frac{P(B_i)\times P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\times P(A/B_i)}$$

**Proof :**




The shaded region represents the event A, A can occur along with $B_1, B_2, B_3, \ldots B_n$ that are mutually exclusive.

∴ $AB_1, AB_2, AB_3, \ldots, AB_n$ are also mutually exclusive.

Also $A = AB_1 \cup AB_2 \cup AB_3 \cup \ldots \cup AB_n$

$$P(A) = P(AB_1) + P(AB_2) + P(AB_3) + \ldots + P(AB_n)$$
$$= \sum_{i=1}^{n} P(AB_i)$$
$$= \sum_{i=1}^{n} P(B_i)\times P(A/B_i) \quad \text{(By conditional probability)}$$

$$P(B_i/A) = \frac{P(B_i)\times P(A/B_i)}{P(A)} = \frac{P(B_i)\times P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\times P(A/B_i)} .$$

**Problem 15:**

Ina bolt factory machines A, B, C manufacture respectively 25%, 35% and 40% of the total. Of their output 5%, 4% and 2% are defective bolts. A bolt is drawn at random from the produce and is found to be defective. What are the probabilities that it was manufactured by machines A, B and C.

**Solution :**

Let $B_1$ be bolt produced by machine A

$B_2$ be bolt produced by machine B

$B_3$ be bolt produced by machine C

Let $A/B_1$ be the defective bolts drawn from machine A

$A/B_2$ be the defective bolts drawn from machine B

$A/B_3$ be the defective bolts drawn from machine C.

$P(B_1) = 0.25$ $\quad$ $P(A/B_1) = 0.05$

$P(B_2) = 0.35$ $\quad$ $P(A/B_2) = 0.04$

$P(B_3) = 0.40$ $\quad$ $P(A/B_3) = 0.02$

Let $B_1/A$ be defective bolts manufactured by machine A

$B_2/A$ be defective bolts manufactured by machine B

$B_3/A$ be defective bolts manufactured by machine C

$P(A) = \sum_{i=1}^{3} P(B_i)\times P(A/B_i) = (0.25) \times (0.05) + (0.35) \times (0.04) + (0.4) \times (0.02)$

$= 0.0345$

$$P(B_1/A) = \frac{P(B_1)\times P(A/B_1)}{P(A)} = 0.3623$$

$$P(B_2/A) = \frac{P(B_2)\times P(A/B_2)}{P(A)} = 0.405$$

$$P(B_3/A) = \frac{P(B_3)\times P(A/B_3)}{P(A)} = 0.231 .$$

**Problem 16 :**

The first bag contains 3 white balls, 2 red balls and 4 black balls. Second bag contains 2 white, 3 red and 5 black balls and third bag contains 3 white, 4 red and 2 black balls. One bag is chosen at random and from it 3 balls are drawn. Out of three balls two balls are white and one is red. What are the probabilities that they were taken from first bag, second bag and third bag.

**Solution :**

Let P(selecting the bag) = $P(A_i) = 1/3$ , $i = 1, 2, 3$.

$P(A/B_1) = \frac{3C_2\,2C_1}{9C_3} = \frac{6}{84}$ $\quad$ $P(A) = \sum_{i=1}^{3} P(B_i)\times P(A/B_i) = 0.0746$

$P(A/B_2) = \frac{2C_2\,3C_1}{10C_3} = \frac{3}{120}$

$$P(A/B_3) = \frac{3C_2\,4C_1}{9C_3} = \frac{12}{84}$$

$$P(B_1/A) = \frac{P(B_1) \times P(A/B_1)}{P(A)} = 0.319$$

$$P(B_2/A) = \frac{P(B_2) \times P(A/B_2)}{P(A)} = 0.4285$$

$$P(B_3/A) = \frac{P(B_3) \times P(A/B_3)}{P(A)} = 0.638$$

# Random Variable

**Random Variable:**

A random variable is a real valued function whose domain is the sample space of a random experiment taking values on the real line $\mathbb{R}$.

**Discrete Random Variable:**

A discrete random variable is one which can take only finite or countable number of values with definite probabilities associated with each one of them.

**Probability mass function:**

Let X be discrete random variable which assuming values $x_1, x_2, \ldots, x_n$ with each of the values, we associate a number called the probability $P(X = x_i) = p(x_i), (i = 1, 2, \ldots, n)$ this is called the probability of $x_i$ satisfying the following conditions

i. $p_i \geq 0 \ \forall i$ i.e., $p_i$'s are all non-negative

ii. $\sum_{i=1}^{n} p_i = p_1 + p_2 + \ldots + p_n = 1$ i.e., the total probability is one.

**Continuous random variable:**

A continuous random variable is one which can assume every value between two specified values with a definite probability associated with each.

**Probability Density Function:**

A function $f$ is said to be the probability density function of a continuous random variable X if it satisfies the following properties.

i. $f(x) \geq 0; \ -\infty < x < \infty$

ii. $\int_{-\infty}^{\infty} f(x) dx = 1.$

**Distribution Function or Cumulative Distribution Function**

i. Discrete Variable:

A distribution function of a discrete random variable X is defined as $P(X \leq x) = \sum_{x_i \leq x} P(x_i)$.

ii. Continuous Variable:

A distribution function of a continuous random variable X is defined as $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x) dx$.

**Mathematical Expectation**

The expected value of the random variable X is defined as

i. If X is discrete random variable $E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$ where $p(x)$ is the probability function of $x$.

ii. If X is continuous random variable $E(X)=\int_{-\infty}^{\infty} xf(x)dx$ where $f(x)$ is the probability density function of $x$.

Properties of Expectation:

1. If $C$ is constant then $E(C)=C$

Proof:

Let $X$ be a discrete random variable then $E(x)=\sum xp(x)$

Now $E(C)=\sum Cp(x)$

$$=C\sum p(x) \qquad \text{since } \sum_{i=1}^{n} p_i = p_1 + p_2 + \ldots + p_n = 1$$

$$=C$$

2. If $a,b$ are constants then $E(ax+b)=aE(x)+b$

Proof:

Let $X$ be a discrete random variable then $E(x)=\sum xp(x)$

Now $E(ax+b)=\sum (ax+b)\,p(x)$

$$=\sum axp(x)+\sum bp(x)$$

$$=a\sum xp(x)+b\sum p(x) \qquad \text{since } \sum_{i=1}^{n} p_i = p_1 + p_2 + \ldots + p_n = 1$$

$$=aE(x)+b$$

3. If $a$ and $b$ are constants then $Var(ax+b)=a^2Var(x)$

Proof:

$$Var(ax+b)=E\left[\left(ax+b-E(ax+b)\right)^2\right]$$

$$=E\left[\left(ax+b-aE(x)-b\right)^2\right]$$

$$=E\left[a^2\left(x-E(x)\right)^2\right]$$

$$=a^2E\left[\left(x-E(x)\right)^2\right]$$

$$=a^2Var(x).$$

4. If $a$ is constant then $Var(ax)=a^2Var(x)$

Proof:

$$Var(ax)=E\left[\left(ax-E(ax)\right)^2\right]$$

$$=E\left[\left(ax-aE(x)\right)^2\right]$$

$$=E\left[a^2\left(x-E(x)\right)^2\right]$$

$$=a^2E\left[\left(x-E(x)\right)^2\right]$$

$$=a^2Var(x).$$

5. Prove that $Var(x) = E(x^2) - [E(x)]^2$

Proof:

$$
\begin{aligned}
Var(x) &= E\left[\left(x - E(x)\right)^2\right] \\
&= E\left[x^2 + \left(E(x)\right)^2 - 2xE(x)\right] \\
&= E\left[x^2 + \mu^2 - 2x\mu\right] \\
&= E\left(x^2\right) + E\left(\mu^2\right) - E(2x\mu) \\
&= E\left(x^2\right) + \mu^2 - 2\mu E(x) \\
&= E\left(x^2\right) + \mu^2 - 2\mu^2 \\
&= E\left(x^2\right) - \mu^2
\end{aligned}
$$

$$Var(x) = E\left(x^2\right) - \left[E(x)\right]^2$$

**Problem.1**

If the probability distribution of $X$ is given as

| | | | | | |
|---|---|---|---|---|---|
| $X$ | : | 1 | 2 | 3 | 4 |
| $P\ X$ | : | 0.4 | 0.3 | 0.2 | 0.1 |

Find $P(1/2 < X < 7/2 / X > 1)$

**Solution:**

$$P\{1/2 < X < 7/2 / X > 1\} = \frac{P\{(1/2 < X < 7/2) \cap X > 1\}}{P(X > 1)}$$

$$= \frac{P(X = 2\, or\, 3)}{P(X = 2, 3\, or\, 4)}$$

$$= \frac{P(X = 2) + P(X = 3)}{P(X = 2) + P(X = 3) + P(X = 4)}$$

$$= \frac{0.3 + 0.2}{0.3 + 0.2 + 0.1} = \frac{0.5}{0.6} = \frac{5}{6}.$$

**Problem.2**

A random variable $X$ has the following probability distribution

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X$ | : | $-2$ | $-1$ | 0 | 1 | 2 | 3 |
| $P\ X$ | : | 0.1 | $K$ | 0.2 | $2K$ | 0.3 | $3K$ |

a) Find $K$, b) Evaluate $P(X < 2)$ and $P(-2 < X < 2)$

b) Find the cdf of $X$ and d) Evaluate the mean of $X$.

**Solution:**

a) Since $\sum P(X) = 1$

$$0.1 + K + 0.2 + 2K + 0.3 + 3K = 1$$

$$6K + 0.6 = 1$$

$$6K = 0.4$$

$$K = \frac{0.4}{6} = \frac{1}{15}$$

b) $P(X < 2) = P(X = -2, -1, 0\, or\, 1)$

$$= P(X = -2) + P(X = -1) + P(X = 0) + P(X = 1)$$

$$= \frac{1}{10} + \frac{1}{15} + \frac{1}{5} + \frac{2}{15}$$

$$= \frac{3 + 2 + 6 + 4}{30} = \frac{15}{30} = \frac{1}{2}$$

$P(-2 < X < 2) = P(X = -1, 0\, or\, 1)$

$$= P(X = -1) + P(X = 0) + P(X = 1)$$

$$=\frac{1}{15}+\frac{1}{5}+\frac{2}{15}$$
$$=\frac{1+3+2}{15}=\frac{6}{15}=\frac{2}{5}$$

c) The distribution function of $X$ is given by $F(x)$ defined by

| $X=x$ | $P(X=x)$ | $F(x)=P(X\le x)$ |
|---|---|---|
| -2 | $\frac{1}{10}$ | $F(x)=P(X\le -2)=\frac{1}{10}$ |
| -1 | $\frac{1}{15}$ | $F(x)=P(X\le -1)=\frac{1}{6}$ |
| 0 | $\frac{2}{10}$ | $F(x)=P(X\le 0)=\frac{11}{30}$ |
| 1 | $\frac{2}{15}$ | $F(x)=P(X\le 1)=\frac{1}{2}$ |
| 2 | $\frac{3}{10}$ | $F(x)=P(X\le 2)=\frac{4}{5}$ |
| 3 | $\frac{3}{15}$ | $F(x)=P(X\le 3)=1$ |

d) Mean of $X$ is defined by $E(X)=\sum xP(x)$

$$E(X)=\left(-2\times\frac{1}{10}\right)+\left(-1\times\frac{1}{15}\right)+\left(0\times\frac{1}{5}\right)+\left(1\times\frac{2}{15}\right)+\left(2\times\frac{3}{10}\right)+\left(3\times\frac{1}{5}\right)$$
$$=-\frac{1}{5}-\frac{1}{15}+\frac{2}{15}+\frac{3}{5}+\frac{3}{5}=\frac{16}{15}.$$

**Problem.3**

A random variable $X$ has the following probability function:

| $X$ | : | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| $P\ X$ | : | 0 | $K$ | $2K$ | $2K$ | $3K$ | $K^2$ | $2K^2$ | $7K^2+K$ |

Find (i) $K$, (ii) Evaluate $P(X<6), P(X\ge 6)$ and $P(0<X<5)$

(iii). Determine the distribution function of $X$.

(iv). $P(1.5<X<4.5/X>2)$

(v). $E(3x-4), Var(3x-4)$

(vi). The smallest value of $n$ for which $P(X\le n)>\frac{1}{2}$.

**Solution:**

**(i)** Since $\sum_{x=0}^{7}P(X)=1,$

$$K+2K+2K+3K+K^2+2K^2+7K^2+K=1$$
$$10K^2+9K-1=0$$
$$K=\frac{1}{10}\quad or\quad K=-1$$

As $P(X)$ cannot be negative $K=\frac{1}{10}$

**(ii)** $P(X<6)=P(X=0)+P(X=1)+...+P(X=5)$

$$=\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100}+...=\frac{81}{100}$$

Now $P(X\geq 6)=1-P(X<6)$

$$=1-\frac{81}{100}=\frac{19}{100}$$

Now $P(0<X<5)=P(X=1)+P(X=2)+P(X=3)=P(X=4)$

$$=K+2K+2K+3K$$

$$=8K=\frac{8}{10}=\frac{4}{5}.$$

**(iii)** The distribution of X is given by $F(x)=P(X\leq x)$

| $X=x$ | $P(X=x)$ | $F(x)=P(X\leq x)$ |
|---|---|---|
| 0 | 0 | $F(x)=P(X\leq 0)=0$ |
| 1 | $\frac{1}{10}$ | $F(x)=P(X\leq 1)=\frac{1}{10}$ |
| 2 | $\frac{2}{10}$ | $F(x)=P(X\leq 2)=\frac{3}{10}$ |
| 3 | $\frac{2}{10}$ | $F(x)=P(X\leq 3)=\frac{5}{10}$ |
| 4 | $\frac{3}{10}$ | $F(x)=P(X\leq 4)=\frac{8}{10}$ |
| 5 | $\frac{1}{100}$ | $F(x)=P(X\leq 5)=\frac{81}{100}$ |
| 6 | $\frac{2}{100}$ | $F(x)=P(X\leq 6)=\frac{83}{100}$ |
| 7 | $\frac{17}{100}$ | $F(x)=P(X\leq 7)=1$ |

**(iv)** $P(1.5<X<4.5/X>2)=\dfrac{P(x=3)+P(x=4)}{1-\left[P(x=0)+P(x=1)+P(x=2)\right]}$

$$=\frac{\frac{5}{10}}{1-\left[\frac{3}{10}\right]}=\frac{5}{7}$$

**(v)** $E(x)=\sum xp(x)$

$$=1\times\frac{1}{10}+2\times\frac{2}{10}+3\times\frac{2}{10}+4\times\frac{3}{10}+5\times\frac{1}{100}+6\times\frac{2}{100}+7\times\frac{17}{100}$$

$E(x)=3.66$

$E(x^2)=\sum x^2 p(x)$

$$=1^2\times\frac{1}{10}+2^2\times\frac{2}{10}+3^2\times\frac{2}{10}+4^2\times\frac{3}{10}+5^2\times\frac{1}{100}+6^2\times\frac{2}{100}+7^2\times\frac{17}{100}$$

$$E\left(x^2\right)=16.8$$

Mean $=E(x)=3.66$

Variance $=E\left(x^2\right)-\left[E(x)\right]^2$

$$=16.8-(3.66)^2$$

$$=3.404$$

**(vi)** The smallest value of $n$ for which $P(X\le n)>\frac{1}{2}$ is 4

**Problem.4**

The probability mass function of random variable $X$ is defined as $P(X=0)=3C^2$, $P(X=1)=4C-10C^2$, $P(X=2)=5C-1$, where $C>0$, and $P(X=r)=0$ if $r\ne 0,1,2$.

Find (i). The value of $C$.

(ii). $P(0<X<2/x>0)$.

(iii). The distribution function of $X$.

(iv). The largest value of $x$ for which $F(x)<\frac{1}{2}$.

**Solution:**

**(i)** Since $\sum_{x=0}^{x=2} p(x)=1$

$$p(0)+p(1)+p(2)=1$$

$$3C^2+4C-10C^2+5C-1=1$$

$$7C^2-9C+2=0$$

$$C=1,\frac{2}{7}$$

$C=1$ is not applicable

$$\therefore C=\frac{2}{7}$$

The Probability distribution is

| | | | | |
|---|---|---|---|---|
| $X$ | : | 0 | 1 | 2 |
| $P(X)$ | : | $\frac{12}{49}$ | $\frac{16}{49}$ | $\frac{21}{49}$ |

(ii) $$P\left[0<x<2\big/x>0\right]=\frac{P\left[(0<x<2)\cap x>0\right]}{P[x>0]}$$

$$=\frac{P[0<x<2]}{P[x>0]}=\frac{P[x=1]}{P[x=1]+P[X=2]}$$

$$P\left[0<x<2\big/x>0\right]=\frac{\frac{16}{49}}{\frac{16}{49}+\frac{21}{49}}=\frac{16}{37}$$

**(iii).** The distribution function of $X$ is

| $X$ | $F(X=x)=P(X\le x)$ |
|---|---|
| 0 | $F(0)=P(X\le 0)=\frac{12}{49}=0.24$ |
| 1 | $F(1)=P(X\le 1)=P(X=0)+P(X=1)=\frac{12}{49}+\frac{16}{49}=0.57$ |
| 2 | $F(2)=P(X\le 2)=P(X=0)+P(X=1)+P(X=2)=\frac{12}{49}+\frac{16}{49}+\frac{21}{49}=1$ |

**(iv)** The Largest value of $x$ for which $F(x)=P(X\le x)<\frac{1}{2}$ is 0.

**Problem.5**

If $P(x)=\begin{cases}\frac{x}{15}; x=1,2,3,4,5\\ 0 \quad ; elsewhere\end{cases}$

Find (i) $P\{X=1 or 2\}$ and (ii) $P\{1/2<X<5/2/x>1\}$

**Solution:**

i) $P(X=1\, or\, 2)=P(X=1)+P(X=2)$

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

ii) $P\left(\frac{1}{2}<X<\frac{5}{2}/x>1\right)=\frac{P\left\{\left(\frac{1}{2}<X<\frac{5}{2}\right)\cap(X>1)\right\}}{P(X>1)}$

$$=\frac{P\{(X=1 or 2)\cap(X>1)\}}{P(X>1)}$$

$$=\frac{P(X=2)}{1-P(X=1)}$$

$$=\frac{2/15}{1-(1/15)}=\frac{2/15}{14/15}=\frac{2}{14}=\frac{1}{7}.$$

**Problem.6**

A continuous random variable $X$ has a probability density function $f(x)=3x^2$, $0\le x\le 1$. Find $'a'$ such that $P(X\le a)=P(X>a)$.

**Solution:**

Since $P(X\le a)=P(X>a)$, each must be equal to $\frac{1}{2}$ because the probability is always 1.

$$\therefore P(X\le a)=\frac{1}{2}$$

$$\Rightarrow \int_0^a f(x)dx=\frac{1}{2}$$

$$\int_0^a 3x^2dx=\frac{1}{2}\Rightarrow 3\left[\frac{x^3}{3}\right]_0^a=a^3=\frac{1}{2}.$$

$$\therefore a = \left(\frac{1}{2}\right)^{\frac{1}{3}}$$

**Problem.7**

A random variable $X$ has the p.d.f $f(x)$ given by $f(x) = \begin{cases} Cxe^{-x}; \; if \; x > 0 \\ 0 \quad ; \; if \; x \le 0 \end{cases}$ Find the value of $C$ and cumulative density function of $X$ .

**Solution:**

Since $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_0^{\infty} Cxe^{-x}dx = 1$$

$$C\left[x\left(-e^{-x}\right) - \left(e^{-x}\right)\right]_0^{\infty} = 1$$

$$C = 1$$

$$\therefore f(x) = \begin{cases} xe^{-x}; x > 0 \\ 0 \quad ; x \le 0 \end{cases}$$

Cumulative Distribution of $x$ is

$$F(x) = \int_0^x f(x)dt = \int_0^x xe^{-x}dx = \left[-xe^{-x} - e^{-x}\right]_0^x = -xe^{-x} - e^{-x} + 1$$

$$= 1-(1+x)e^{-x},\ x > 0.$$

**Problem.8**

If a random variable $X$ has the p.d.f $f(x) = \begin{cases} \frac{1}{2}(x+1); -1 < x < 1 \\ \quad 0 \quad ; otherwise \end{cases}$. Find the mean and variance of $X$ .

**Solution:**

$$\text{Mean} = \mu_1' = \int_{-1}^{1} xf(x)\,dx = \frac{1}{2}\int_{-1}^{1} x(x+1)\,dx = \frac{1}{2}\int_{-1}^{1}\left(x^2+x\right)dx$$

$$= \frac{1}{2}\left(\frac{x^3}{3} + \frac{x^2}{2}\right)_{-1}^{1} = \frac{1}{3}$$

$$\mu_2' = \int_{-1}^{1} x^2 f(x)dx = \frac{1}{2}\int_{-1}^{1}\left(x^3+x^2\right)dx = \frac{1}{2}\left[\frac{x^4}{4} + \frac{x^3}{3}\right]_{-1}^{1}$$

$$= \frac{1}{2}\left[\frac{1}{4} + \frac{1}{3} - \frac{1}{4} + \frac{1}{3}\right]$$

$$= \frac{1}{2}.\frac{2}{3} = \frac{1}{3}$$

$$Variance = \mu_2' - \left(\mu_1'\right)^2$$

$$= \frac{1}{3} - \frac{1}{9} = \frac{3-1}{9} = \frac{2}{9}.$$

**Problem.9**

A continuous random variable X that can assume any value between $X = 2$ and $X = 5$ has a probability density function given by $f(x)=k(1+x)$ . Find $P(X < 4)$.

**Solution:**

Given X is a continuous random variable whose pdf is $f(x) = \begin{cases} k(1+x), 2 < x < 5 \\ 0 \ \ , Otherwise \end{cases}$.

Since $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{2}^{5} k(1+x)dx = 1$

$$k\left[\frac{(1+x)^2}{2}\right]_2^5 = 1$$

$$k\left[\frac{(1+5)^2}{2} - \frac{(1+2)^2}{2}\right] = 1$$

$$k\left[18 - \frac{9}{2}\right] = 1$$

$$k\left[\frac{27}{2}\right] = 1 \Rightarrow k = \frac{2}{27}$$

$$\therefore f(x) = \begin{cases} \frac{2(1+x)}{27}, 2 < x < 5 \\ 0 \ \ , Otherwise \end{cases}$$

$$P(X < 4) = \frac{2}{27}\int_{2}^{4} (1+x)\,dx$$

$$= \frac{2}{27}\left[\frac{(1+x)^2}{2}\right]_2^4 = \frac{2}{27}\left[\frac{(1+4)^2}{2} - \frac{(1+2)^2}{2}\right] = \frac{2}{27}\left[\frac{25}{2} - \frac{9}{2}\right] = \frac{2}{27}\frac{16}{2} = \frac{16}{27}.$$

**Problem.10**

A random variable $X$ has density function given by $f(x) = \begin{cases} 2e^{-2x}; x \geq 0 \\ 0 \quad ; x < 0 \end{cases}$. Find m.g.f

**Solution:**

$$M_X(t) = E\left(e^{tx}\right) = \int_0^{\infty} e^{tx} f(x)\,dx = \int_0^{\infty} e^{tx} 2e^{-2x} dx$$

$$= 2\int_0^{\infty} e^{(t-2)x} dx$$

$$= 2\left[\frac{e^{(t-2)x}}{t-2}\right]_0^{\infty} = \frac{2}{2-t}, t < 2.$$

**Problem.11**

The pdf of a random variable X is given by $f(x) = \begin{cases} 2x, \ 0 \leq x \leq b \\ 0, \ \ otherwise \end{cases}$. For what value of b is $f(x)$ a valid pdf? Also find the cdf of the random variable X with the above pdf.

**Solution:**

Given $f(x)=\begin{cases}2x, & 0\le x\le b\\ 0, & otherwise\end{cases}$

Since $\int_{-\infty}^{\infty} f(x)\,dx=1 \Rightarrow \int_{0}^{b} 2x\,dx=1$

$$\left[2\frac{x^2}{2}\right]_0^b=1$$

$$\left[b^2-0\right]=1 \Rightarrow \text{b} = 1$$

$$\therefore f(x)=\begin{cases}2x, & 0\le x\le 1\\ 0, & otherwise\end{cases}$$

$$F(x)=P(X\le x)=\int_0^x f(x)\,dx=\int_0^x 2x dx=\left[2\frac{x^2}{2}\right]_0^x=x^2\ ,\ 0\le x\le 1$$

$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{x} 0\,dx=0\ ,\ x<0$$

$$F(x)=P(X\le x)=\int_{-\infty}^{0} f(x)\,dx+\int_{0}^{1} f(x)\,dx+\int_{1}^{x} f(x)\,dx$$

$$=\int_{-\infty}^{0} 0\ dx+\int_{0}^{1} 2x\,dx+\int_{1}^{x} 0\,dx\ =\left[2\frac{x^2}{2}\right]_0^1=1\ ,\ x>1$$

$$F(x)=\begin{cases}0, & x<0\\ x^2, & 0\le x\le 1\\ 1, & x>1\end{cases}$$

**Problem.12**

A random variable $X$ has density function $f(x)=\begin{cases}\dfrac{K}{1+x^2}, & -\infty<x<\infty\\ 0\ , & Otherwise\end{cases}$. Determine $K$ and the distribution functions. Evaluate the probability $P(x\ge 0)$.

**Solution:**

Since $\int_{-\infty}^{\infty} f(x)dx=1$

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

$$K\int_{\infty}^{\infty}\frac{dx}{1+x^2}=1$$

$$K\left(\tan^{-1}x\right)_{-\infty}^{\infty}=1$$

$$K\left(\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)\right)=1$$

$$K\pi=1$$

$$K=\frac{1}{\pi}$$

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{x}\frac{K}{1+x^2}dx$$

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

$$F(x)=\frac{1}{\pi}\left[\frac{\pi}{2}+\tan^{-1}x\right],-\infty<x<\infty$$

$$P(X\geq 0)=\frac{1}{\pi}\int_0^{\infty}\frac{dx}{1+x^2}=\frac{1}{\pi}\left(\tan^{-1}x\right)_0^{\infty}$$

$$=\frac{1}{\pi}\left(\frac{\pi}{2}-\tan^{-1}0\right)=\frac{1}{2}.$$

**Problem.13**

If $X$ has the probability density function $f(x)=\begin{cases}Ke^{-3x}, x>0\\ 0 \quad ,otherwise\end{cases}$ find $K$, $P[0.5\leq X\leq 1]$ and the mean of $X$ .

**Solution:**

Since $\int_{-\infty}^{\infty} f(x)dx=1$

$$\int_0^{\infty} Ke^{-3x}dx=1$$

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

$$\frac{K}{3}=1$$

$$K=3$$

$$P(0.5\leq X\leq 1)=\int_{0.5}^{1} f(x)dx=3\int_{0.5}^{1} e^{-3x}dx=\cancel{3}\left(\frac{e^{-3}-e^{-1.5}}{-\cancel{3}}\right)=\left[e^{-1.5}-e^{-3}\right]$$

Mean of $X=E(x)=\int_0^{\infty} xf(x)dx=3\int_0^{\infty} xe^{-3x}dx$

$$=3\left[x\left(\frac{-e^{-3x}}{3}\right)-1\left(\frac{e^{-3x}}{9}\right)\right]_0^{\infty}=\frac{3\times 1}{9}=\frac{1}{3}$$

Hence the mean of $X=E(X)=\frac{1}{3}$.

**Problem.14**

If $X$ is a continuous random variable with pdf given by

$$f(x)=\begin{cases} Kx & in \quad 0\le x\le 2 \\ 2K & in \quad 2\le x\le 4 \\ 6K-Kx & in \quad 4\le x\le 6 \\ 0 & elsewhere \end{cases}$$. Find the value of $K$ and also the cdf $F(x)$.

**Solution:**

Since $\int_{\infty}^{\infty} F(x)\,dx=1$

$$\int_0^2 Kxdx+\int_2^4 2Kdx+\int_4^6 (6k-kx)dx=1$$

$$K\left[\left(\frac{x^2}{2}\right)_0^2+(2x)_2^4+\int_4^6\left(6x-\frac{x^2}{2}\right)_4^6\right]=1$$

$$K\left[\cancel{2}+\cancel{8}-4+36-18-24+8\right]=1$$

$$8K=1$$

$$K=\frac{1}{8}$$

We know that $F(x)=\int_{-\infty}^{x} f(x)\,dx$

If $x<0$, then $F(x)=\int_{-\infty}^{x} f(x)\,dx=0$

If $x\in(0,2)$, then $F(x)=\int_{-\infty}^{x} f(x)\,dx$

$$F(x)=\int_{-\infty}^{0} f(x)\,dx+\int_0^x f(x)\,dx$$

$$=\int_{-\infty}^{0} 0dx+\int_0^x Kxdx=\int_{-\infty}^{0} 0dx+\frac{1}{8}\int_0^x xdx$$

$$F(x)=\left(\frac{x^2}{16}\right)_0^x=\frac{x^2}{16},0\le x\le 2$$

If $x\in(2,4)$, then $F(x)=\int_{-\infty}^{0} f(x)\,dx+\int_0^2 f(x)\,dx+\int_2^x f(x)\,dx$

$$=\int_{-\infty}^{0} 0dx+\int_0^2 Kxdx+\int_2^x 2Kdx$$

$$=\int_0^2\frac{x}{8}dx+\int_2^x\frac{1}{4}dx=\left(\frac{x^2}{16}\right)_0^2+\left(\frac{x}{4}\right)_2^x$$

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

$$F(x)=\frac{x}{4}-\frac{4}{16}=\frac{x-1}{4},2\le x<4$$

If $x \in (4,6)$, then $F(x) = \int_{-\infty}^{0} 0dx + \int_{0}^{2} Kxdx + \int_{2}^{4} 2Kdx + \int_{4}^{x} K(6-x)dx$

$$= \int_{0}^{2} \frac{x}{8} dx + \int_{2}^{4} \frac{1}{4} dx + \int_{4}^{x} \frac{1}{8}(6-x) dx$$

$$= \left(\frac{x^2}{16}\right)_0^2 + \left(\frac{x}{4}\right)_2^4 + \left(\frac{6x}{8} - \frac{x^2}{16}\right)_4^x$$

$$= \frac{1}{4} + 1 - \frac{1}{2} + \frac{6x}{8} - \frac{x^2}{16} - 3 + 1$$

$$= \frac{4 + 16 - 8 + 12x - x^2 - 48 + 16}{16}$$

$$F(x) = \frac{-x^2 + 12x - 20}{16}, 4 \le x \le 6$$

If $x > 6$, then $F(x) = \int_{-\infty}^{0} 0dx + \int_{0}^{2} Kxdx + \int_{2}^{4} 2Kdx + \int_{4}^{6} K(6-x)dx + \int_{6}^{\infty} 0dx$

$$F(x) = 1, x \ge 6$$

$$\therefore F(x) = \begin{cases} 0 & ; x \le 0 \\ \frac{x^2}{16} & ; 0 \le x \le 2 \\ \frac{1}{4}(x-1) & ; 2 \le x \le 4 \\ \frac{-1}{16}(20 - 12x + x^2) & ; 4 \le x \le 6 \\ 1 & ; x \ge 6 \end{cases}$$

**Problem.15**

A random variable $X$ has the P.d.f $f(x) = \begin{cases} 2x, 0 < x < 1 \\ 0 \;\;, Otherwise \end{cases}$

Find (i) $P\left(X < \frac{1}{2}\right)$ (ii) $P\left(\frac{1}{4} < x < \frac{1}{2}\right)$ (iii) $P\left(X > \frac{3}{4} / X > \frac{1}{2}\right)$

**Solution:**

(i) $P\left(x < \frac{1}{2}\right) = \int_{0}^{1/2} f(x)dx = \int_{0}^{1/2} 2xdx = 2\left(\frac{x^2}{2}\right)_0^{1/2} = \frac{2 \times 1}{8} = \frac{1}{4}$

(ii) $P\left(\frac{1}{4} < x < \frac{1}{2}\right) = \int_{1/4}^{1/2} f(x)dx = \int_{1/4}^{1/2} 2xdx = 2\left(\frac{x^2}{2}\right)_{1/4}^{1/2}$

$$= 2\left(\frac{1}{8} - \frac{1}{32}\right) = \left(\frac{1}{4} - \frac{1}{16}\right) = \frac{3}{16}.$$

(iii) $P\left(X > \frac{3}{4} / X > \frac{1}{2}\right) = \frac{P\left(X > \frac{3}{4} \cap X > \frac{1}{2}\right)}{P\left(X > \frac{1}{2}\right)} = \frac{P\left(X > \frac{3}{4}\right)}{P\left(X > \frac{1}{2}\right)}$

$$P\left(X>\frac{3}{4}\right)=\int_{3/4}^{1} f(x)\,dx=\int_{3/4}^{1} 2x\,dx=2\left(\frac{x^2}{2}\right)_{3/4}^{1}=1-\frac{9}{16}=\frac{7}{16}$$

$$P\left(X>\frac{1}{2}\right)=\int_{1/2}^{1} f(x)\,dx=\int_{1/2}^{1} 2x\,dx=2\left(\frac{x^2}{2}\right)_{1/2}^{1}=1-\frac{1}{4}=\frac{3}{4}$$

$$P\left(X>\frac{3}{4}/X>\frac{1}{2}\right)=\frac{\frac{7}{16}}{\frac{3}{4}}=\frac{7}{16}\times\frac{4}{3}=\frac{7}{12}.$$

**Problem.16**

Let the random variable $X$ have the p.d.f $f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}} & ,x>0\\ 0 & ,otherwise.\end{cases}$ .Find the moment generating function, mean & variance of $X$ .

**Solution:**

$$M_X(t)=E\left(e^{tx}\right)=\int_{-\infty}^{\infty} e^{tx} f(x)\,dx=\int_{0}^{\infty} e^{tx}\frac{1}{2}e^{-x/2}dx$$

$$=\frac{1}{2}\int_{0}^{\infty} e^{-\left(\frac{1}{2}-t\right)x}dx=\frac{1}{2}\left[\frac{e^{-\left(\frac{1}{2}-t\right)x}}{-\left(\frac{1}{2}-t\right)}\right]_0^{\infty}=\frac{1}{1-2t},\ if\ t<\frac{1}{2}.$$

$$E(X)=\left[\frac{d}{dt}M_X(t)\right]_{t=0}=\left[\frac{2}{(1-2t)^2}\right]_{t=0}=2$$

$$E\left(X^2\right)=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left[\frac{8}{(1-2t)^3}\right]_{t=0}=8$$

$$Var(X)=E\left(X^2\right)-\left[E(X)\right]^2=8-4=4.$$

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF SCIENCE AND HUMANITIES

DEPARTMENT OF MATHEMATICS

# UNIT – II - Probability Distribution – SMTA1402

# SMTA1402 - Probability and Statistics

## Unit-2 Probability Distribution

## Discrete type

**Binomial distribution:**

A random variable $X$ is said to follow binomial distribution if it assumes only non negative values and its probability mass function is given by
$P(X=x)=p(x)=\begin{cases} nC_x p^x q^{n-x}, x=0,1,2,\ldots,n; q=1-p \\ 0, otherwise \end{cases}$

Notation: $X \sim B(n,p)$ read as $X$ is following binomial distribution with parameter $n$ and $p$.

**Problem.1**

Find m.g.f. of Binomial distribution and find its mean and variance.

**Solution:**

M.G.F.of Binomial distribution:-

$$M_X(t)=E\left[e^{tx}\right]=\sum_{x=0}^{n} e^{tx} P(X=x)$$

$$=\sum_{x=0}^{n} nC_x x\, P^x q^{n-x} e^{tx}$$

$$=\sum_{x=0}^{n} nC_x \left(pe^t\right)^x q^{n-x}$$

$$M_X(t)=\left(q+pe^t\right)^n$$

Mean of Binomial distribution

$$\text{Mean}=E(X)=M_X'(0)$$

$$=\left[n\left(q+pe^t\right)^{n-1} pe^t\right]_{t=0}=np \text{ Since } q+p=1$$

$$E\left(X^2\right)=M_X''(0)$$

$$=\left[n(n-1)\left(q+pe^t\right)^{n-2}\left(pe^t\right)^2+npe^t\left(q+pe^t\right)^{n-1}\right]_{t=0}$$

$$E\left(X^2\right)=n(n-1)p^2+np$$

$$=n^2p^2+np(1-p)=n^2p^2+npq$$

$$\text{Variance}=E\left(X^2\right)-\left[E[X]\right]^2=npq$$

$$Mean=np \,;Variance=npq$$

**Problem.2**

Comment the following: "The mean of a binomial distribution is 3 and variance is 4

**Solution:**

In binomial distribution, mean $>$ variance but Variance $<$ Mean

Since Variance $=4$ &Mean $=3$, the given statement is wrong.

**Problem.3**

If $X$ and $Y$ are independent binomial variates $B\left(5,\frac{1}{2}\right)$ and $B\left(7,\frac{1}{2}\right)$ find $P\left[X+Y=3\right]$

**Solution:**

$X+Y$ is also a binomial variate with parameters $n_1+n_2=12$ & $p=\frac{1}{2}$

$$\therefore P\left[X+Y=3\right]=12C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^9=\frac{55}{2^{10}}$$

**Problem.4**

(i). Six dice are thrown 729 times. How many times do you expect atleast 3 dice show 5 or 6 ?

(ii) Six coins are tossed 6400 times. Using the Poisson distribution, what is the approximate probability of getting six heads 10 times?

**Solution:**

(i). Let $X$ be the number of times the dice shown 5 or 6

$$P\left[5\ or\ 6\right]=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$$

$$\therefore P=\frac{1}{3}\text{ and }q=\frac{2}{3}$$

Here $n=6$

By Binomial theorem,

$$P\left[X=x\right]=6C_x\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{6-x}\text{ where }x=0,1,2...6.$$

$$P\left[X\geq 3\right]=P(3)+P(4)+P(5)+P(6)$$

$$=6C_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3+6C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2+6C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)+6C_6\left(\frac{1}{3}\right)^6$$

$$=0.3196$$

$\therefore$ Expected number of times atleast 3 dies to show 5 or 6 $=N\times P\left[X\geq 3\right]$

$$=729\times 0.3196=233.$$

(ii). Probability of getting six heads in one toss of six coins is $p=\left(\frac{1}{2}\right)^6$,

$$\lambda=np=6400\times\left(\frac{1}{2}\right)^6=100$$

Let $X$ be the number of times getting 6 heads $P(X=10)=\dfrac{e^{-100}(100)^{10}}{10!}=1.025\times 10^{-30}$

**Poisson distribution:**

A random variable $X$ is said to follow Poisson distribution if it assumes only non negative values and its probability mass function is given by

$$P\left(X=x\right)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!}; x=0,1,2,...;\lambda>0\\ 0, otherwise\end{cases}$$

Notation: $X \sim P(\lambda)$ read as $X$ is following Poisson distribution with parameter $\lambda$.

**Poisson distribution as limiting form of binomial distribution:**

Poisson distribution is a limiting case of Binomial distribution under the following conditions:

(i). $n$ the number of trials is indefinitely large, (i.e.) $n \to \infty$

(ii). $p$ the constant probability of success in each trial is very small (i.e.) $p \to 0$

(iii). $np = \lambda$ is finite.

**Proof:**

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

Let $np = \lambda$

$$\therefore \qquad p = \frac{\lambda}{n},\ q = 1 - \frac{\lambda}{n}$$

$$\therefore \qquad p(x) = nc_x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-(x-1))\,\cancel{(n-x)!}}{x!\,\cancel{(n-x)!}}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{1.\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}\,\cancel{n^x}\,\frac{\lambda^x}{\cancel{n^x}}\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$p(x) = 1.\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Taking limit $n \to \infty$ on both sides

$$\lim_{n\to\infty} p(x) = \frac{\lambda^x}{x!}\lim_{n\to\infty}\left[\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\left(1 - \frac{\lambda}{n}\right)^{n-x}\right]$$

$$= \frac{\lambda^x}{x!}\lim_{n\to\infty}\left[\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\right]\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{-x}\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{n}$$

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2,\ldots$$

**Problem.1**

Criticise the following statement: "The mean of a Poisson distribution is 5 while the standard deviation is 4".

**Solution:**

For a Poisson distribution mean and variance are same. Hence this statement is not true.

**Geometric distribution:**

A random variable $X$ is said to have a Geometric distribution if it assumes only non negative values and its probability mass function is given by

$$P(X=x)=\begin{cases} q^{x-1}p; x=1,2,\ldots;0<p\le 1 \\ 0, otherwise \end{cases}$$

**Problem.1**

Find the Moment generating function of geometric distribution and find its Mean and Variance

**Solution:**

$$M_X(t)=E\left(e^{tX}\right)$$

$$=\sum_{x=1}^{\infty} e^{tx} q^{x-1} p$$

$$=\sum_{x=1}^{\infty} pe^t \left(qe^t\right)^{x-1}$$

$$=pe^t\left(1+qe^t+\left(qe^t\right)^2+\cdots\right)$$

$$=pe^t\left(1-qe^t\right)^{-1}$$

$$M_X(t)=\frac{pe^t}{1-qe^t}$$

$$\mu_1'=M_x'(0)=\left[\frac{d}{dt}\left(\frac{pe^t}{\left(1-qe^t\right)}\right)\right]_{t=0}=\left[\left(\frac{pe^t}{\left(1-qe^t\right)^2}\right)\right]_{t=0}=\frac{1}{p}$$

$$\mu_2'=M_x''(0)=\left[\frac{d^2}{dt^2}\left(\frac{pe^t}{\left(1-qe^t\right)}\right)\right]_{t=0}=\left[\frac{d}{dt}\left(\frac{pe^t}{\left(1-qe^t\right)^2}\right)\right]_{t=0}=\frac{1+q}{p^2}$$

$$\text{Mean}=\mu_1'=\frac{1}{p}$$

$$\text{Variance}=\mu_2'-\left(\mu_1'\right)^2=\frac{1+q}{p^2}-\left(\frac{1}{p}\right)^2=\frac{q}{p^2}$$

**Problem.2**

State and prove Memoryless property of geometric distribution.

**Solution:**

If $X$ has a geometric distribution, then for any two positive integer '$s$' and '$t$' $P\left[X>s+t \middle/ X>s\right]=P\left[X>t\right]$.

The p.m.f of the geometric random variable $X$ is $P(X=x)=q^{x-1}p$ , $x=1,2,3,\ldots.$

$$P\left[X>s+t \middle/ X>s\right]=\frac{P\left[X>s+t\cap X>s\right]}{P\left[X>s\right]}=\frac{P\left[X>s+t\right]}{P\left[X>s\right]} \quad ------(1)$$

$$\therefore P\left[X>t\right]=\sum_{x=t+1}^{\infty} q^{x-1}p = q^t p+q^{t+1}p+q^{t+2}p+.... = q^t p\left[1+q+q^2+q^3+....\right]$$

$$= q^t p(1-q)^{-1} = q^t p(p)^{-1} = q^t$$

Hence $P\left[X>s+t\right]=q^{s+t}$ and $P\left[X>s\right]=q^s$

$$(1) \Rightarrow P\left[X>s+t \middle/ X>s\right]=\frac{P\left[X>s+t \cap X>s\right]}{P\left[X>s\right]}=\frac{q^{s+t}}{q^s}=q^t=P[X>t]$$

$$\Rightarrow P\left[X>s+t \middle/ X>s\right]=P\left(X>t\right)$$

**Problem.3**

If the probability is $\frac{1}{4}$ that a man will hit a target what is the chance that he will hit the target for the first time in the 7th trial?

**Solution:**

The required probability is

$$P\left[FFFFFFS\right]=P\left(F\right)P\left(F\right)P\left(F\right)P\left(F\right)P\left(F\right)P\left(F\right)P\left(S\right)$$

$$=q^6 p=\left(\frac{3}{4}\right)^6 .\left(\frac{1}{4}\right)=0.0445.$$

**Problem.4**

A die is cast until 6 appears what is the probability that it must cast more then five times?

**Solution:**

Probability of getting six $=\frac{1}{6}$

$\therefore p=\frac{1}{6}$ & $q=1-\frac{1}{6}$

Let $x$: No of throws for getting the number 6. By geometric distribution $P\left[X=x\right]=q^{x-1}p, x=1,2,3....$

Since 6 can be got either in first, second……throws.

To find $P\left[X\geq 6\right]=1-P\left[X<6\right]$

$$=1-\sum_{x=1}^{5}\left(\frac{5}{6}\right)^{x-1}.\frac{1}{6}$$

$$=1-\left[\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)+\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)\right]$$

$$=1-\frac{\frac{1}{6}\left[1-\left(\frac{5}{6}\right)^5\right]}{1-\frac{5}{6}}=\left(\frac{5}{6}\right)^5=0.4019$$

**Problem.5**

Suppose that a trainee soldier shoots a target an independent fashion. If the probability that the target is shot on any one shot is 0.8.

(i) What is the probability that the target would be hit on 6th attempt?

(ii) What is the probability that it takes him less than 5 shots?

**Solution:**

Here $p = 0.8, q = 1 - p = 0.2$

$$P[X = x] = q^{x-1}p, x = 1, 2...$$

(i) The probability that the target would be hit on the 6th attempt $= P[X = 6]$

$$= (0.2)^5 (0.8) = 0.00026$$

(ii) The probability that it takes him less than 5 shots $= P[X < 5]$

$$= \sum_{x=1}^{4} q^{x-1} p = 0.8 \sum_{x=1}^{4} (0.2)^{x-1}$$

$$= 0.8[1 + 0.2 + 0.04 + 0.008] = 0.9984$$

**Continuous type**

**Uniform (or) Rectangular distribution:**

A continuous random variable $X$ is said to have a uniform distribution over an interval $(a,b)$ if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, a < x < b \\ 0, otherwise \end{cases}$$

**Problem.1**

If $X$ is uniformly distributed with Mean $1$ and Variance $\frac{4}{3}$, find $P[X > 0]$

**Solution:**

If $X$ is uniformly distributed over $(a,b)$, then

$$E(X) = \frac{b+a}{2} \text{ and } V(X) = \frac{(b-a)^2}{12}$$

$$\therefore \frac{b+a}{2} = 1 \Rightarrow a + b = 2$$

$$\Rightarrow \frac{(b-a)^2}{12} = \frac{4}{3} \Rightarrow (b-a)^2 = 16$$

$\Rightarrow a + b = 2$ & $b - a = 4$ We get $b = 3, a = -1$

$\therefore a = -1$ & $b = 3$ and probability density function of $x$ is

$$f(x) = \begin{cases} \frac{1}{4}; -1 < x < 3 \\ 0; Otherwise \end{cases}$$

$$P[x < 0] = \int_{-1}^{0} \frac{1}{4} dx = \frac{1}{4}[x]_{-1}^{0} = \frac{1}{4}.$$

**Exponential distribution:**

A continuous random variable $X$ assuming non negative values is said to have an exponential distribution with parameter $\theta > 0$ , if its probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, otherwise \end{cases}$$

**Problem.1**

Find the moment generating function of Exponential distribution and find its mean and variance.

**Solution:**

We know that $f(x)=\begin{cases}\lambda e^{-\lambda x}, x\ge 0\\ 0, otherwise\end{cases}$

$$M_X(t)=E\left(e^{tx}\right)=\int_0^\infty e^{tx}f(x)dx=\int_0^\infty \lambda e^{-\lambda x}e^{tx}dx$$

$$=\lambda\int_0^\infty e^{-x(\lambda-t)}dx$$

$$=\lambda\left[\frac{e^{-x(\lambda-t)}}{-(\lambda-t)}\right]_0^\infty=\frac{\lambda}{\lambda-t}$$

$$\text{Mean}=\mu_1'=\left[\frac{d}{dt}M_X(t)\right]_{t=0}=\left[\frac{\lambda}{(\lambda-t)^2}\right]_{t=0}=\frac{1}{\lambda}$$

$$\mu_2'=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left[\frac{\lambda(2)}{(\lambda-t)^3}\right]_{t=0}=\frac{2}{\lambda^2}$$

$$\text{Variance}=\mu_2'-\left(\mu_1'\right)^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}.$$

**Problem.2**

State and prove the memoryless property of exponential distribution.

**Solution:**

Statement:

If $X$ is exponentially distributed with parameters $\lambda$, then for any two positive integers‘s’ and‘t’, $P[x>s+t/x>s]=P[x>t]$

Proof:

The p.d.f of X is $f(x)=\begin{cases}\lambda e^{-\lambda x}, x\ge 0\\ 0 \quad ,Otherwise\end{cases}$

$$\therefore P[X>k]=\int_k^\infty \lambda e^{-\lambda x}dx=\left[-e^{-\lambda x}\right]_k^\infty=e^{-\lambda k}$$

$$\therefore P[X>s+t/x>s]=\frac{P[x>s+t\cap x>s]}{P[x>s]}$$

$$=\frac{P[X>s+t]}{P[X>s]}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}$$

$$=P[x>t]$$

**Problem.3**

A component has an exponential time to failure distribution with mean of 10,000 hours.

(i). The component has already been in operation for its mean life. What is the probability that it will fail by 15,000 hours?

(ii). At 15,000 hours the component is still in operation. What is the probability that it will operate for another 5000 hours.

**Solution:**

Let $X$ denote the time to failure of the component then $X$ has exponential distribution with $Mean = 1000$ hours.

$$\therefore \frac{1}{\lambda} = 10,000 \Rightarrow \lambda = \frac{1}{10,000}$$

The p.d.f. of $X$ is $f(x) = \begin{cases} \frac{1}{10,000} e^{-\frac{x}{10,000}} & , x \geq 0 \\ 0 & , otherwise \end{cases}$

(i) Probability that the component will fail by 15,000 hours given it has already been in operation for its mean life $= P\left[x < 15,000 / x > 10,000\right]$

$$= \frac{P\left[10,000 < X < 15,000\right]}{P\left[X > 10,000\right]}$$

$$= \frac{\int_{10,000}^{15,000} f(x)\,dx}{\int_{10,000}^{\infty} f(x)\,dx} = \frac{e^{-1} - e^{-1.5}}{e^{-1}}$$

$$= \frac{0.3679 - 0.2231}{0.3679} = 0.3936.$$

(ii) Probability that the component will operate for another 5000 hours given that it is in operational 15,000 hours $= P\left[X > 20,000 / X > 15,000\right]$

$= P\left[x > 5000\right]$ [By memoryless prop]

$$= \int_{5000}^{\infty} f(x)\,dx = e^{-0.5} = 0.6065$$

**Normal distribution:**

A random variable $X$ is said to have a Normal distribution with parameters $\mu$ (mean) and $\sigma^2$ (variance) if its probability density function is given by the probability law

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

Notation: $X \sim N\left(\mu, \sigma^2\right)$ read as $X$ is following normal distribution with mean $\mu$ and variance $\sigma^2$ are called parameter.

**Problem.1**

Prove that "For standard normal distribution $N(0,1)$, $M_X(t) = e^{\frac{t^2}{2}}$.

**Solution:**

Moment generating function of Normal distribution

$$= M_X(t) = E\left[e^{tx}\right]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Put $z = \frac{x-\mu}{\sigma}$ then $\sigma dz = dx, -\infty < Z < \infty$

$$\therefore M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z + \mu) - \frac{z^2}{2}} dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{z^2}{2} - t\sigma z\right)} dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2 + \left(\frac{\sigma^2 t^2}{2}\right)} dz$$

$$= \frac{e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2} dz$$

$\because$ the total area under normal curve is unity, we have $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2} dz = 1$

Hence $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$ $\therefore$ For standard normal variable $N(0,1)$

$$M_X(t) = e^{\frac{t^2}{2}}$$

**Problem.2**

State and prove the additive property of normal distribution.

**Solution:**

Statement:

If $X_1, X_2, \ldots, X_n$ are $n$ independent normal random variates with mean $\left(\mu_1, \sigma_1^{\,2}\right)$, $\left(\mu_2, \sigma_2^{\,2}\right), \ldots \left(\mu_n, \sigma_n^{\,2}\right)$ then $X_1 + X_2 + \ldots + X_n$ also a normal random variable with mean $\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^{\,2}\right)$.

Proof:

We know that. $M_{X_1+X_2+\ldots+X_n}(t) = M_{X_1}(t) M_{X_2}(t) \ldots M_{X_n}(t)$

But $M_{X_i}(t) = e^{\mu_i t + \frac{t^2 \sigma_i^2}{2}}$, $i = 1, 2 \ldots n$

$$M_{X_1+X_2+\ldots+X_n}(t) = e^{\mu_1 t + \frac{t^2 \sigma_1^2}{2}} e^{\mu_2 t + \frac{t^2 \sigma_2^2}{2}} \ldots e^{\mu_n t + \frac{t^2 \sigma_n^2}{2}}$$

$$= e^{(\mu_1 + \mu_2 + \ldots + \mu_n)t + \frac{\left(\sigma_1^{\,2} + \sigma_2^{\,2} + \ldots + \sigma_n^{\,2}\right)t^2}{2}}$$

$$= e^{\sum_{i=1}^{n} \mu_i t + \frac{\sum_{i=1}^{n} \sigma_i^2 t^2}{2}}$$

By uniqueness MGF, $X_1 + X_2 + \ldots + X_n$ follows normal random variable with parameter $\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^{\,2}\right)$.

This proves the property.

**Problem.3**

$X$ is a normal variate with $mean = 30$ and $S.D = 5$ Find the following $P\left[26 \le X \le 40\right]$

**Solution:**

$X \sim N\left(30, 5^2\right)$

$\therefore \mu = 30 \;\&\; \sigma = 5$

Let $Z = \dfrac{X - \mu}{\sigma}$ be the standard normal variate

$$P\left[26 \le X \le 40\right] = P\left[\frac{26-30}{5} \le Z \le \frac{40-30}{5}\right]$$

$$= P\left[-0.8 \le Z \le 2\right] = P\left[-0.8 \le Z \le 0\right] + P\left[0 \le Z \le 2\right]$$

$$= P\left[0 \le Z\; 0.8\right] + \left[0 \le z \le 2\right]$$

$$= 0.2881 + 0.4772 = 0.7653.$$

**Problem.4**

The average percentage of marks of candidates in an examination is 45 will a standard deviation of 10 the minimum for a pass is 50%.If 1000 candidates appear for the examination, how many can be expected marks. If it is required, that double that number should pass, what should be the average percentage of marks?

**Solution:**

Let $X$ be marks of the candidates

Then $X \sim N\left(42, 10^2\right)$

Let $z = \dfrac{X - 42}{10}$

$P\left[X > 50\right] = P\left[Z > 0.8\right]$

$$= 0.5 - P[0 < z < 0.8]$$
$$= 0.5 - 0.2881 = 0.2119$$

Since 1000 students write the test, nearly 212 students would pass the examination.

If double that number should pass, then the no of passes should be 424.

We have to find $z_1$, such that $P[Z > z_1] = 0.424$

$$\therefore P[0 < z < z_1] = 0.5 - 0.424 = 0.076$$

From tables, $z = 0.19$

$$\therefore z_1 = \frac{50 - x_1}{10} \Rightarrow x_1 = 50 - 10z_1$$
$$= 50 - 1.9 = 48.1$$

The average mark should be 48 nearly.

**Problem.5**

Given that $X$ is normally distribution with mean 10 and probability $P[X > 12] = 0.1587$. What is the probability that $X$ will fall in the interval $(9,11)$.

**Solution:**

Given $X$ is normally distributed with mean $\mu = 10$.

Let $z = \dfrac{x - \mu}{\sigma}$ be the standard normal variate.

For $X = 12, z = \dfrac{12 - 10}{\sigma} \Rightarrow z = \dfrac{2}{\sigma}$

Put $z_1 = \dfrac{2}{\sigma}$

Then $P[X > 12] = 0.1587$

$$P[Z > Z_1] = 0.1587$$
$$\therefore 0.5 - p[0 < z < z_1] = 0.1587$$
$$\Rightarrow P[0 < z < z_1] = 0.3413$$

From area table $P[0 < z < 1] = 0.3413$

$$\therefore Z_1 = 1 \Rightarrow \frac{2}{\sigma} = 1$$

To find $P[9 < x < 11]$

For $X = 9,\ z = -\dfrac{1}{2}$ and $X = 11,\ z = \dfrac{1}{2}$

$$\therefore P[9 < X < 11] = P[-0.5 < z < 0.5]$$
$$= 2P[0 < z < 0.5]$$
$$= 2 \times 0.1915 = 0.3830$$

31. In a normal distribution 31% of the items are under 45 and 8% are over 64.Find the mean and standard deviation of the distribution.

Solution:

Let $\mu$ be the mean and $\sigma$ be the standard deviation.

Then $P[X \le 45] = 0.31$ and $P[X \ge 64] = 0.08$

When $X = 45$, $Z = \frac{45-\mu}{\sigma} = -z_1$

$\therefore z_1$ is the value of $z$ corresponding to the area $\int_0^{z_1} \phi(z)dz = 0.19$

$\therefore z_1 = 0.495$

$45 - \mu = -0.495\sigma$ ---(1)

When $X = 64$, $Z = \frac{64-\mu}{\sigma} = z_2$

$\therefore z_2$ is the value of $z$ corresponding to the area $\int_0^{z_2} \phi(z)dz = 0.42$

$\therefore z_2 = 1.405$

$64 - \mu = 1.405\sigma$ ---(2)

Solving (1) & (2) We get $\mu = 10$ (approx) & $\sigma = 50$ (approx)

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

**SCHOOL OF SCIENCE AND HUMANITIES**

**DEPARTMENT OF MATHEMATICS**

# UNIT – III - Two Dimensional Random Variable – SMTA1402

# Sathyabama Institute of Science and Technology

## SMTA1402 - Probability and Statistics

### UNIT-3 TWO DIMENSIONAL RANDOM VARIABLE

1. Let $X$ and $Y$ have joint density function $f(x,y)=2,\ 0<x<y<1$. Find the marginal density function. Find the conditional density function $Y$ given $X=x$.

Solution:

Marginal density function of $X$ is given by

$$f_X(x)=f(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

$$=\int_x^1 f(x,y)\,dy=\int_x^1 2dy=2(y)_x^1$$

$$=2(1-x),0<x<1.$$

Marginal density function of $Y$ is given by

$$f_Y(y)=f(y)=\int_{-\infty}^{\infty} f(x,y)\,dx$$

$$=\int_0^y 2dx=2y,0<y<1\ .$$

Conditional distribution function of $Y$ given $X=x$ is $f\left(y/x\right)=\dfrac{f(x,y)}{f(x)}=\dfrac{2}{2(1-x)}=\dfrac{1}{1-x}$.

2. Verify that the following is a distribution function. $F(x)=\begin{cases}0 & ,x<-a\\ \frac{1}{2}\left(\frac{x}{a}+1\right) & ,-a<x<a\\ 1 & ,x>a\end{cases}$.

Solution:

$F(x)$ is a distribution function only if $f(x)$ is a density function.

$$f(x)=\frac{d}{dx}\left[F(x)\right]=\frac{1}{2a},\ -a<x<a$$

$$\int_{-\infty}^{\infty} f(x)=1$$

$$\therefore \int_{-a}^{a}\frac{1}{2a}dx=\frac{1}{2a}\left[x\right]_{-a}^{a}=\frac{1}{2a}\left[a-(-a)\right]$$

$$=\frac{1}{2a}.2a=1\ .$$

Therefore, it is a distribution function.

3. Prove that $\int_{x_1}^{x_2} f_X(x)\,dx = p(x_1 < x < x_2)$

Solution:

$$\int_{x_1}^{x_2} f_X(x)\,dx = \left[F_X(x)\right]_{x_1}^{x_2}$$
$$= F_X(x_2) - F_X(x_1)$$
$$= P[X \le x_2] - P[X \le x_1]$$
$$= P[x_1 \le X \le x_2]$$

4. A continuous random variable $X$ has a probability density function $f(x) = 3x^2,\ \ 0 \le x \le 1$. Find '$a$' such that $P(X \le a) = P(X > a)$.

Solution:

Since $P(X \le a) = P(X > a)$, each must be equal to $\frac{1}{2}$ because the probability is always 1.

$$\therefore P(X \le a) = \frac{1}{2}$$

$$\Rightarrow \int_0^a f(x)\,dx = \frac{1}{2}$$

$$\int_0^a 3x^2 dx = \frac{1}{2} \Rightarrow 3\left[\frac{x^3}{3}\right]_0^a = a^3 = \frac{1}{2}.$$

$$\therefore a = \left(\frac{1}{2}\right)^{\frac{1}{3}}$$

5. Suppose that the joint density function $f(x,y) = \begin{cases} Ae^{-x-y}, & 0 \le x \le y,\ 0 \le y \le \infty \\ 0 & , otherwise \end{cases}$ Determine $A$.

Solution:

Since $f(x,y)$ is a joint density function

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1.$$

$$\Rightarrow \int_0^{\infty}\int_0^{y} Ae^{-x}e^{-y}\,dxdy = 1$$

$$\Rightarrow A\int_0^{\infty} e^{-y}\left(\frac{e^{-x}}{-1}\right)_0^y dy = 1$$

$$\Rightarrow A\int_0^{\infty}\left[e^{-y} - e^{-2y}\right]dy = 1$$

$$\Rightarrow A\left[\frac{e^{-y}}{-1}-\frac{e^{-2y}}{-2}\right]_0^\infty=1$$

$$\Rightarrow A\left[\frac{1}{2}\right]=1\Rightarrow A=2$$

6. Examine whether the variables $X$ and $Y$ are independent, whose joint density function is $f\left(x,y\right)=xe^{-x(y+1)}, 0<x,y<\infty$.

Solution:

The marginal probability function of $X$ is

$$f_X\left(x\right)=f\left(x\right)=\int_{-\infty}^{\infty}f\left(x,y\right)dy=\int_0^\infty xe^{-x(y+1)}dy$$

$$=x\left[\frac{e^{-x(y+1)}}{-x}\right]_0^\infty=-\left[0-e^{-x}\right]=e^{-x},$$

The marginal probability function of $Y$ is

$$f_Y\left(y\right)=f\left(y\right)=\int_{-\infty}^{\infty}f\left(x,y\right)dx=\int_0^\infty x\,e^{-x(y+1)}dx$$

$$=x\left\{\left[\frac{e^{-x(y+1)}}{-(y+1)}\right]_0^\infty-\left[\frac{e^{-x(y+1)}}{(y+1)^2}\right]\right\}_0^\infty$$

$$=\frac{1}{(y+1)^2}$$

Here $f\left(x\right).f\left(y\right)=e^{-x}\times\frac{1}{(1+y)^2}\neq f\left(x,y\right)$

$\therefore$ $X$ and $Y$ are not independent.

7. If $X$ has an exponential distribution with parameter 1. Find the pdf of $y=\sqrt{x}$

Solution:

Since $y=\sqrt{x}$, $x=y^2$

Since $X$ has an exponential distribution with parameter 1, the pdf of $X$ is given by

$$f_X\left(x\right)=e^{-x},\ x>0 \qquad \left[\quad f\left(x\right)=\lambda e^{-\lambda x},\lambda=1\right]$$

$$\therefore f_Y\left(y\right)=f_X\left(x\right)\left|\frac{dx}{dy}\right|$$

$$=e^{-x}2y=2ye^{-y^2}$$

$$f_Y\left(y\right)=2ye^{-y^2},\ \ y>0$$

8. If $X$ is uniformly distributed random variable in $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$, Find the probability density function of $Y=tanX$.

Solution:

Given $Y = tanX \Rightarrow x = tan^{-1} y$

$$\therefore \frac{dx}{dy} = \frac{1}{1+y^2}$$

Since $X$ is uniformly distribution in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$,

$$f_X(x) = \frac{1}{b-a} = \frac{1}{\frac{\pi}{2}+\frac{\pi}{2}}$$

$$f_X(x) = \frac{1}{\pi}, \ -\frac{\pi}{2} < x < \frac{\pi}{2}$$

Now $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{\pi}\left(\frac{1}{1+y^2}\right), -\infty < y < \infty$

$$\therefore f_Y(y) = \frac{1}{\pi(1+y^2)}, -\infty < y < \infty$$

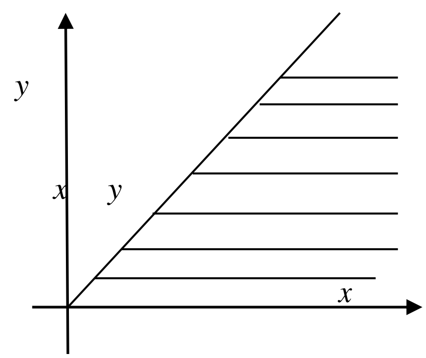
9. If the Joint probability density function of $(x, y)$ is given by $f(x,y) = 24y(1-x)$, $0 \le y \le x \le 1$ Find $E(XY)$.

Solution:

$$E(xy) = \int_0^1 \int_y^1 xyf(x,y)\,dxdy$$

$$= 24\int_0^1 \int_y^1 xy^2(1-x)\,dxdy$$

$$= 24\int_0^1 y^2\left[\frac{1}{6} - \frac{y^2}{2} + \frac{y^3}{3}\right]dy = \frac{4}{15}.$$




10. If $X$ and $Y$ are random Variables, Prove that $Cov(X,Y) = E(XY) - E(X)E(Y)$

Solution:

$$cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

$$= E\left(XY - \bar{X}Y - \bar{Y}X + \bar{X}\bar{Y}\right)$$

$$= E(XY) - \bar{X}E(Y) - \bar{Y}E(X) + \bar{X}\bar{Y}$$

$$= E(XY) - \bar{X}\bar{Y} - \bar{X}\bar{Y} + \bar{X}\bar{Y}$$

$$= E(XY) - E(X)E(Y) \qquad \left[\ E(X) = \bar{X}, E(Y) = \bar{Y}\right]$$

11. If $X$ and $Y$ are independent random variables prove that $cov(x,y) = 0$

Proof:

$cov(x, y) = E(xy) - E(x)E(y)$

But if $X$ and $Y$ are independent then $E(xy) = E(x)E(y)$

$cov(x, y) = E(x)E(y) - E(x)E(y)$

$cov(x, y) = 0.$

12. Write any two properties of regression coefficients.

Solution:

1. Correction coefficients is the geometric mean of regression coefficients
2. If one of the regression coefficients is greater than unity then the other should be less than 1.

$$b_{xy} = r\frac{\sigma_y}{\sigma_x} \text{ and } b_{yx} = r\frac{\sigma_x}{\sigma_y}$$

If $b_{xy} > 1$ then $b_{yx} < 1$.

13. Write the angle between the regression lines.

The slopes of the regression lines are

$$m_1 = r\frac{\sigma_y}{\sigma_x}, m_2 = \frac{1}{r}\frac{\sigma_y}{\sigma_x}$$

If $\theta$ is the angle between the lines, Then

$$\tan\theta = \frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\left[\frac{1-r^2}{r}\right]$$

When $r = 0$, that is when there is no correlation between x and y, $\tan\theta = \infty$ (or) $\theta = \frac{\pi}{2}$ and so the regression lines are perpendicular

When $r = 1$ or $r = -1$, that is when there is a perfect correlation $+ve$ or $-ve$, $\theta = 0$ and so the lines coincide.

15. i). Two random variables are said to be orthogonal if correction is <u>zero.</u>

ii). If $X = Y$ then correlation coefficients between them is <u>1.</u>

16.a). The joint probability density function of a bivariate random variable $X, Y$ is $f_{XY}(x,y)=\begin{cases} k(x+y), \ 0<x<2, \ 0<y<2 \\ 0 \quad , \ otherwise \end{cases}$ where $'k'$ is a constant.

i. Find $k$.
ii. Find the marginal density function of $X$ and $Y$.
iii. Are $X$ and $Y$ independent?
iv. Find $f_{Y/X}\left( y/x \right)$ and $f_{X/Y}\left( x/y \right)$.

Solution:

(i). Given the joint probability density function of a brivate random variable $(X,Y)$ is

$$f_{XY}(x,y)=\begin{cases} K(x+y), \ 0<x<2, \ 0<y<2 \\ 0 \quad , \ otherwise \end{cases}$$

Here $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy=1 \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(x+y)\,dxdy=1$

$$\int_0^2\int_0^2 K(x+y)\,dxdy=1 \Rightarrow K\int_0^2\left[\frac{x^2}{2}+xy\right]_0^2 dy=1$$

$$\Rightarrow K\int_0^2 (2+2y)\,dy=1$$

$$\Rightarrow K\left[2y+y^2\right]_0^2=1$$

$$\Rightarrow K\left[8-0\right]=1$$

$$\Rightarrow K=\frac{1}{8}$$

(ii). The marginal p.d.f of $X$ is given by

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\frac{1}{8}\int_0^2 (x+y)\,dy$$

$$=\frac{1}{8}\left[xy+\frac{y^2}{2}\right]_0^2=\frac{1+x}{4}$$

$\therefore$ The marginal p.d.f of $X$ is

$$f_X(x)=\begin{cases} \dfrac{x+1}{4}, \ 0<x<2 \\ 0 \quad , \ otherwise \end{cases}$$

The marginal p.d.f of $Y$ is

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\frac{1}{8}\int_0^2 (x+y)\,dx$$

$$=\frac{1}{8}\left[\frac{x^2}{2}+yx\right]_0^2$$

$$=\frac{1}{8}\left[2+2y\right]=\frac{y+1}{4}$$

$\therefore$ The marginal p.d.f of $Y$ is

$$f_Y(y)=\begin{cases}\frac{y+1}{4}, & 0<y<2\\ 0 & , otherwise\end{cases}$$

(iii). To check whether $X$ and $Y$ are independent or not.

$$f_X(x)\,f_Y(y)=\frac{(x+1)}{4}\frac{(y+1)}{4}\neq f_{XY}(x,y)$$

Hence $X$ and $Y$ are not independent.

(iv). Conditional p.d.f $f_{Y/X}\left(y/x\right)$ is given by

$$f_{Y/X}\left(y/x\right)=\frac{f(x,y)}{f_X(x)}=\frac{\frac{1}{8}(x+y)}{\frac{1}{4}(x+1)}=\frac{1}{2}\frac{(x+y)}{(x+1)}$$

$$f_{Y/X}\left(y/x\right)=\frac{1}{2}\left(\frac{x+y}{x+1}\right),\ 0<x<2,\ 0<y<2$$

(v) $P\left(0<y<\frac{1}{2}\Big/x=1\right)=\int_0^2 f_{Y/X}\left(y/x=1\right)dy$

$$=\frac{1}{2}\int_0^{\frac{1}{2}}\frac{1+y}{2}\,dy=\frac{5}{32}.$$

17.a). If $X$ and $Y$ are two random variables having joint probability density function $f(x,y)=\begin{cases}\frac{1}{8}(6-x-y), & 0<x<2,\ 2<y<4\\ 0 & , otherwise\end{cases}$ Find (i) $P(X<1\cap Y<3)$

(ii) $P(X+Y<3)$ (iii) $P\left(X<1\big/Y<3\right)$.

b). Three balls are drawn at random without replacement from a box containing 2 white, 3 red and 4 black balls. If $X$ denotes the number of white balls drawn and $Y$ denotes the number of red balls drawn find the joint probability distribution of $X,Y$ .

Solution:

a).

$$P(X<1\cap Y<3)=\int_{y=-\infty}^{y=3}\int_{x=-\infty}^{x=1} f(x,y)\,dxdy$$

$$=\int_{y=2}^{y=3}\int_{x=0}^{x=1}\frac{1}{8}(6-x-y)\,dxdy$$

$$= \frac{1}{8}\int_{2}^{3}\int_{0}^{1}(6-x-y)\,dxdy$$

$$= \frac{1}{8}\int_{2}^{3}\left[6x-\frac{x^2}{2}-xy\right]_0^1 dy$$

$$= \frac{1}{8}\int_{2}^{3}\left[\frac{11}{2}-y\right]dy = \frac{1}{8}\left[\frac{11y}{2}-\frac{y^2}{2}\right]_2^3$$

$$P(X<1\cap Y<3) = \frac{3}{8}$$

(ii). $P(X+Y<3) = \int_{0}^{1}\int_{2}^{3-x}\frac{1}{8}(6-x-y)\,dydx$

$$= \frac{1}{8}\int_{0}^{1}\left[6y-xy-\frac{y^2}{2}\right]_2^{3-x} dx$$

$$= \frac{1}{8}\int_{0}^{1}\left[6(3-x)-x(3-x)-\frac{(3-x)^2}{2}-[12-2x-2]\right]dx$$

$$= \frac{1}{8}\int_{0}^{1}\left[18-6x-3x+x^2-\frac{(9+x^2-6x)}{2}-(10-2x)\right]dx$$

$$= \frac{1}{8}\int_{0}^{1}\left[18-9x+x^2-\frac{9}{2}-\frac{x^2}{2}+\frac{6x}{2}-10+2x\right]dx$$

$$= \frac{1}{8}\int_{0}^{1}\left[\frac{7}{2}-4x+\frac{x^2}{2}\right]dx$$

$$= \frac{1}{8}\left[\frac{7x}{2}-\frac{4x^2}{2}+\frac{x^3}{6}\right]_0^1 = \frac{1}{8}\left[\frac{7}{2}-2+\frac{1}{6}\right]$$

$$= \frac{1}{8}\left[\frac{21-12+1}{6}\right] = \frac{1}{8}\left(\frac{10}{6}\right) = \frac{5}{24}.$$

(iii). $P\left(X<1 \big/ Y<3\right) = \dfrac{P(x<1\cap y<3)}{P(y<3)}$

The Marginal density function of Y is $f_Y(y) = \int_{0}^{2} f(x,y)\,dx$

$$= \int_{0}^{2}\frac{1}{8}(6-x-y)\,dx$$

$$= \frac{1}{8}\left[6x-\frac{x^2}{2}-yx\right]_0^2$$

$$= \frac{1}{8}[12-2-2y]$$

$$= \frac{5-y}{4},\ 2 < y < 4.$$

$$P\left(X < 1 \middle/ Y < 3\right) = \frac{\int_{x=0}^{x=1}\int_{y=2}^{y=3} \frac{1}{8}(6-x-y)\,dxdy}{\int_{y=2}^{y=3} f_Y(y)\,dy}$$

$$= \frac{\frac{3}{8}}{\int_2^3 \left(\frac{5-y}{4}\right)dy} = \frac{\frac{3}{8}}{\frac{1}{4}\left[5y - \frac{y^2}{2}\right]_2^3}$$

$$= \frac{3}{8} \times \frac{8}{5} = \frac{3}{5}.$$

b). Let $X$ takes 0, 1, 2 and $Y$ takes 0, 1, 2 and 3.

$P(X = 0, Y = 0) = P($drawing 3 balls none of which is white or red)

$P($ all the 3 balls drawn are black)

$$= \frac{4C_3}{9C_3} = \frac{4\times3\times2\times1}{9\times8\times7} = \frac{1}{21}.$$

$P(X = 0, Y = 1) = P($drawing 1 red ball and 2 black balls)

$$= \frac{3C_1 \times 4C_2}{9C_3} = \frac{3}{14}$$

$P(X = 0, Y = 2) = P($ drawing 2 red balls and 1 black ball)

$$= \frac{3C_2 \times 4C_1}{9C_3} = \frac{3\times2\times4\times3}{9\times8\times7} = \frac{1}{7}.$$

$P(X = 0, Y = 3) = P($ all the three balls drawn are red and no white ball)

$$= \frac{3C_3}{9C_3} = \frac{1}{84}$$

$P(X = 1, Y = 0) = P($drawing 1White and no red ball)

$$= \frac{2C_1 \times 4C_2}{9C_3} = \frac{\frac{2\times4\times3}{1\times2}}{\frac{9\times8\times7}{1\times2\times3}}$$

$$= \frac{12\times1\times2\times3}{9\times8\times7} = \frac{1}{7}.$$

$P(X = 1, Y = 1) = P($drawing 1White and 1 red ball)

$$= \frac{2C_1 \times 3C_1}{9C_3} = \frac{2\times3}{\frac{9\times8\times7}{1\times2\times3}} = \frac{2}{7}$$

$P(X=1,Y=2)=P($ drawing 1White and 2 red ball)

$$=\frac{2C_1\times 3C_2}{9C_3}=\frac{2\times 3\times 2}{\frac{9\times 8\times 7}{1\times 2\times 3}}=\frac{1}{14}$$

$P(X=1,Y=3)=0$ (Since only three balls are drawn)

$P(X=2,Y=0)=P($ drawing 2 white balls and no red balls)

$$=\frac{2C_2\times 4C_1}{9C_3}=\frac{1}{21}$$

$P(X=2,Y=1)=P($ drawing 2 white balls and no red balls)

$$=\frac{2C_2\times 3C_1}{9C_3}=\frac{1}{28}$$

$P(X=2,Y=2)=0$

$P(X=2,Y=3)=0$

The joint probability distribution of $X,Y$ may be represented as

| X \ Y | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | $\frac{1}{21}$ | $\frac{3}{14}$ | $\frac{1}{7}$ | $\frac{1}{84}$ |
| 1 | $\frac{1}{7}$ | $\frac{2}{7}$ | $\frac{1}{14}$ | 0 |
| 2 | $\frac{1}{21}$ | $\frac{1}{28}$ | 0 | 0 |

18.a). Two fair dice are tossed simultaneously. Let $X$ denotes the number on the first die and $Y$ denotes the number on the second die. Find the following probabilities.

(i) $P(X+Y)=8$, (ii) $P(X+Y\geq 8)$, (iii) $P(X=Y)$ and (iv) $P\left(X+Y=6/_{Y=4}\right)$.

b) The joint probability mass function of a bivariate discrete random variable $(X,Y)$ in given by the table.

| Y \ X | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.1 | 0.1 | 0.2 |
| 2 | 0.2 | 0.3 | 0.1 |

Find

i. The marginal probability mass function of $X$ and $Y$.

ii. The conditional distribution of $X$ given $Y=1$.

iii. $P(X+Y<4)$

Solution:

a). Two fair dice are thrown simultaneously

$$S=\begin{Bmatrix}(1,1)(1,2)...(1,6)\\(2,1)(2,2)...(2,6)\\ . \quad . \quad ... \quad .\\(6,1)(6,2)...(6,6)\end{Bmatrix},\ n(S)=36$$

Let $X$ denotes the number on the first die and $Y$ denotes the number on the second die.

Joint probability density function of $(X,Y)$ is $P(X=x,Y=y)=\frac{1}{36}$ for

$x=1,2,3,4,5,6$ and $y=1,2,3,4,5,6$

(i) $X+Y=\{$ the events that the no is equal to 8 $\}$

$$=\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$

$$P(X+Y=8)=P(X=2,Y=6)+P(X=3,Y=5)+P(X=4,Y=4)$$
$$+P(X=5,Y=3)+P(X=6,Y=2)$$
$$=\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}=\frac{5}{36}$$

(ii) $P(X+Y\geq 8)$

$$X+Y=\begin{Bmatrix}(2,6)\\(3,5),(3,6)\\(4,4),(4,5),(4,6)\\(5,3),(5,4)(5,5),(5,6)\\(6,2),(6,3),(6,4),(6,5)(6,6)\end{Bmatrix}$$

$$\therefore P(X+Y\geq 8)=P(X+Y=8)+P(X+Y=9)+P(X+Y=10)$$
$$+P(X+Y=11)+P(X+Y=12)$$
$$=\frac{5}{36}+\frac{4}{36}+\frac{3}{36}+\frac{2}{36}+\frac{1}{36}=\frac{15}{36}=\frac{5}{12}$$

(iii) $P(X=Y)$

$$P(X=Y)=P(X=1,Y=1)+P(X=2,Y=2)+......+P(X=6,Y=6)$$
$$=\frac{1}{36}+\frac{1}{36}+..........+\frac{1}{36}=\frac{6}{36}=\frac{1}{6}$$

(iv) $P\left(X+Y=6 \middle/ Y=4\right)=\dfrac{P(X+Y=6\cap Y=4)}{P(Y=4)}$

Now $P(X+Y=6\cap Y=4)=\frac{1}{36}$

$P(Y=4)=\frac{6}{36}$

$$\therefore P\left(X+Y=6 \big/ Y=4\right)=\frac{\frac{1}{36}}{\frac{6}{36}}=\frac{1}{6}.$$

b). The joint probability mass function of $(X,Y)$ is

| $Y \backslash X$ | 1 | 2 | 3 | Total |
|---|---|---|---|---|
| 1 | 0.1 | 0.1 | 0.2 | 0.4 |
| 2 | 0.2 | 0.3 | 0.1 | 0.6 |
| Total | 0.3 | 0.4 | 0.3 | 1 |

From the definition of marginal probability function

$$P_X(x_i)=\sum_{y_j} P_{XY}(x_i, y_j)$$

When $X=1$,

$$P_X(x_i)=P_{XY}(1,1)+P_{XY}(1,2)$$
$$=0.1+0.2=0.3$$

When $X=2$,

$$P_X(x=2)=P_{XY}(2,1)+P_{XY}(2,2)$$
$$=0.2+0.3=0.4$$

When $X=3$,

$$P_X(x=3)=P_{XY}(3,1)+P_{XY}(3,2)$$
$$=0.2+0.1=0.3$$

$\therefore$ The marginal probability mass function of $X$ is

$$P_X(x)=\begin{cases}0.3 & when\ x=1\\ 0.4 & when\ x=2\\ 0.3 & when\ x=3\end{cases}$$

The marginal probability mass function of $Y$ is given by $P_Y(y_j)=\sum_{x_i} P_{XY}(x_i, y_j)$

When $Y \quad 1$, $P_Y(y=1)=\sum_{x_i=1}^{3} P_{XY}(x_i,1)$

$$=P_{XY}(1,1)+P_{XY}(2,1)+P_{XY}(3,1)$$
$$=0.1+0.1+0.2=0.4$$

When $Y \quad 2$, $P_Y(y=2)=\sum_{x_i=1}^{3} P_{XY}(x_i,2)$

$$=P_{XY}(1,2)+P_{XY}(2,2)+P_{XY}(3,2)$$
$$=0.2+0.3+0.1=0.6$$

$\therefore$ Marginal probability mass function of $Y$ is

$$P_Y(y)=\begin{cases}0.4 & when\ y=1\\ 0.6 & when\ y=2\end{cases}$$

(ii) The conditional distribution of $X$ given $Y \quad 1$ is given by

$$P\left(X = x \middle/ Y = 1\right) = \frac{P\left(X = x \cap Y = 1\right)}{P\left(Y = 1\right)}$$

From the probability mass function of $Y$, $P \quad y \quad 1 \qquad P_y \; 1 \qquad 0.4$

When $X \quad 1$, $P\left(X = 1 \middle/ Y = 1\right) = \dfrac{P\left(X = 1 \cap Y = 1\right)}{P\left(Y = 1\right)}$

$$= \frac{P_{XY}(1,1)}{P_Y(1)} = \frac{0.1}{0.4} = 0.25$$

When $X \quad 2$, $P\left(X = 2 \middle/ Y = 1\right) = \dfrac{P_{XY}(2,1)}{P_Y(1)} = \dfrac{0.1}{0.4} = 0.25$

When $X \quad 3$, $P\left(X = 3 \middle/ Y = 1\right) = \dfrac{P_{XY}(3,1)}{P_Y(1)} = \dfrac{0.2}{0.4} = 0.5$

(iii). $P\left(X + Y < 4\right) = P\left\{(x, y) / x + y < 4 \;\; Where \;\; x = 1,2,3; \;\; y = 1,2\right\}$

$$= P\left\{(1,1),(1,2),(2,1)\right\}$$

$$= P_{XY}(1,1) + P_{XY}(1,2) + P_{XY}(2,1)$$

$$= 0.1 + 0.1 + 0.2 = 0.4$$

19.a). If $X$ and $Y$ are two random variables having the joint density function $f(x, y) = \dfrac{1}{27}(x + 2y)$ where $x$ and $y$ can assume only integer values 0, 1 and 2, find the conditional distribution of $Y$ for $X \quad x$.

b). The joint probability density function of $X, Y$ is given by $f_{XY}(x, y) = \begin{cases} xy^2 + \dfrac{x^2}{8}, & 0 \le x \le 2, \quad 0 \le y \le 1 \\ 0 & , \; otherwise \end{cases}$. Find (i) $P \;\; X \quad 1$, (ii) $P \;\; X \quad Y$ and (iii) $P \;\; X \quad Y \quad 1$

Solution:

a). Given $X$ and $Y$ are two random variables having the joint density function

$$f(x, y) = \frac{1}{27}(x + 2y) - - - - (1)$$

Where $x \quad 0,1,2$ and $y \quad 0,1,2$

Then the joint probability distribution $X$ and $Y$ becomes as follows

| $X \backslash Y$ | 0 | 1 | 2 | $f_1(x)$ |
|---|---|---|---|---|
| 0 | 0 | $\frac{1}{27}$ | $\frac{2}{27}$ | $\frac{3}{27}$ |

| | | | | |
|---|---|---|---|---|
| 1 | $\frac{2}{27}$ | $\frac{3}{27}$ | $\frac{4}{27}$ | $\frac{9}{27}$ |
| 2 | $\frac{4}{27}$ | $\frac{5}{27}$ | $\frac{6}{27}$ | $\frac{15}{27}$ |

The marginal probability distribution of $X$ is given by $f_1(X)=\sum_j P(x,y)$ and is calculated in the above column of above table.

The conditional distribution of $Y$ for $X$ is given by $f_1\left(Y=y/X=x\right)=\frac{f(x,y)}{f_1(x)}$ and is obtained in the following table.

| Y \ X | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0 | $\frac{1}{3}$ | $\frac{2}{3}$ |
| 1 | $\frac{1}{9}$ | $\frac{3}{9}$ | $\frac{5}{9}$ |
| 2 | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{2}$ |

$$P\left(Y=0/X=0\right)=\frac{P(X=0,Y=0)}{P(X=0)}=\frac{0}{\frac{6}{27}}=0$$

$$P\left(Y=1/X=0\right)=\frac{P(X=0,Y=1)}{P(X=0)}=\frac{\frac{2}{27}}{\frac{6}{27}}=\frac{1}{3}$$

$$P\left(Y=2/X=0\right)=\frac{P(X=0,Y=2)}{P(X=0)}=\frac{\frac{4}{27}}{\frac{6}{27}}=\frac{2}{3}$$

$$P\left(Y=0/X=1\right)=\frac{P(X=1,Y=0)}{P(X=1)}=\frac{\frac{1}{27}}{\frac{9}{27}}=\frac{1}{9}$$

$$P\left(Y=1/X=1\right)=\frac{P(X=1,Y=1)}{P(X=1)}=\frac{\frac{3}{27}}{\frac{9}{27}}=\frac{3}{9}=\frac{1}{3}$$

$$P\left(Y=2/X=1\right)=\frac{P(X=1,Y=2)}{P(X=1)}=\frac{\frac{5}{27}}{\frac{9}{27}}=\frac{5}{9}$$

$$P\left(Y=0/X=2\right)=\frac{P\left(X=2,Y=0\right)}{P\left(X=2\right)}=\frac{\frac{2}{27}}{\frac{12}{27}}=\frac{1}{6}$$

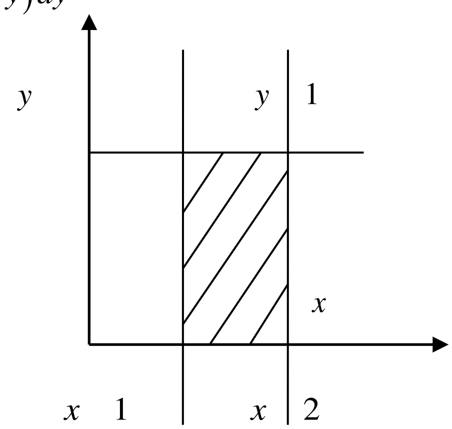
$$P\left(Y=1/X=2\right)=\frac{P\left(X=2,Y=1\right)}{P\left(X=2\right)}=\frac{\frac{4}{27}}{\frac{12}{27}}=\frac{1}{3}$$

$$P\left(Y=2/X=2\right)=\frac{P\left(X=2,Y=2\right)}{P\left(X=2\right)}=\frac{\frac{6}{27}}{\frac{12}{27}}=\frac{1}{2}$$

b). Given the joint probability density function of $X,Y$ is $f_{XY}\left(x+y\right)=xy^2+\frac{x^2}{8}$, $0\le x\le 2,\ 0\le y\le 1$

(i). $P\left(X>1\right)=\int_1^\infty f_X\left(x\right)dx$

The Marginal density function of $X$ is $f_X\left(x\right)=\int_0^1 f\left(x,y\right)dy$

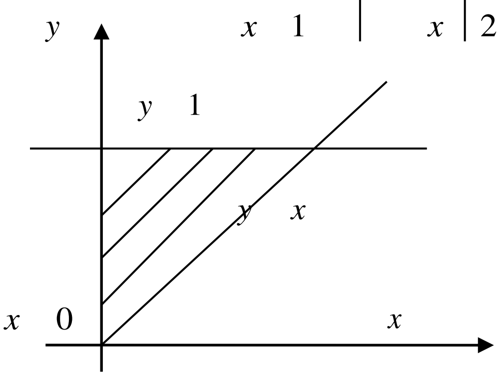



$$f_X\left(x\right)=\int_0^1\left(xy^2+\frac{x^2}{8}\right)dy$$

$$=\left[\frac{xy^2}{3}+\frac{x^2y}{8}\right]_0^1=\frac{x}{3}+\frac{x^2}{8},\ 1<x<2$$

$$P\left(X>1\right)=\int_1^2\left(\frac{x}{3}+\frac{x^2}{8}\right)dx$$

$$=\left[\frac{x^2}{6}+\frac{x^3}{24}\right]_1^2=\frac{19}{24}.$$




(ii) $P\left(X<Y\right)=\int\int_{R_2} f_{XY}\left(x,y\right)dxdy$

$$P\left(X<Y\right)=\int_{y=0}^{1}\int_{x=0}^{y}\left(xy^2+\frac{x^2}{8}\right)dxdy$$

$$=\int_0^1\left[\frac{x^2y^2}{2}+\frac{x^3}{24}\right]_0^y dy$$

$$=\int_0^1\left(\frac{y^4}{2}+\frac{y^3}{24}\right)dy=\left[\frac{y^5}{10}+\frac{y^4}{96}\right]_0^1$$

$$=\frac{1}{10}+\frac{1}{96}=\frac{96+10}{960}=\frac{53}{480}$$

(iii) $P\left(X+Y\le 1\right)=\iint\limits_{R_3} f_{XY}\left(x,y\right)dxdy$

Where $R_3$ is the region

$$P\left(X+Y\le 1\right)=\int_{y=0}^{1}\int_{x=0}^{1-y}\left(xy^2+\frac{x^2}{8}\right)dxdy$$

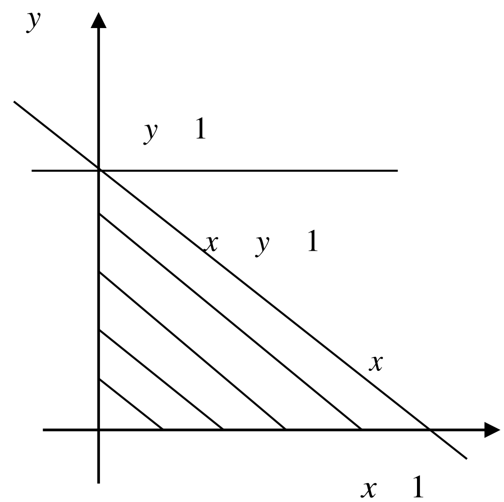
$$=\int_{y=0}^{1}\left[\left(\frac{x^2y^2}{2}+\frac{x^3}{24}\right)\right]_0^{1-y}dy$$

$$=\int_{y=0}^{1}\left(\frac{\left(1-y\right)^2y^2}{2}+\frac{\left(1-y\right)^3}{24}\right)dy$$

$$=\int_0^1\left(\frac{\left(1+y^2-2y\right)y^2}{2}+\frac{\left(1-y\right)^3}{24}\right)dy$$

$$=\left[\left[\frac{y^3}{3}+\frac{y^5}{5}-\frac{2y^2}{4}\right]\frac{1}{2}+\frac{\left(1-y\right)^4}{96}\right]_0^1$$

$$=\frac{1}{6}+\frac{1}{10}-\frac{1}{4}+\frac{1}{96}=\frac{13}{480}.$$




20).a). If the joint distribution functions of $X$ and $Y$ is given by

$$F\left(x,y\right)=\begin{cases}\left(1-e^{x}\right)\left(1-e^{-y}\right), & x>0,\ y>0\\ 0 & ,\ otherwise\end{cases}$$

i. Find the marginal density of $X$ and $Y$.
ii. Are $X$ and $Y$ independent.
iii. $P\ 1\quad X\quad 3,\ 1\quad Y\quad 2$ .

b). The joint probability distribution of $X$ and $Y$ is given by

$$f\left(x,y\right)=\begin{cases}\dfrac{6-x-y}{8}, & 0<x<2,\ 2<y<4\\ 0 & ,\ otherwise\end{cases}$$ . Find $P\ 1\quad Y\quad 3\!\!\diagup\!\! X\quad 2$ .

Solution:

a). Given $F\left(x,y\right)=\left(1-e^{-x}\right)\left(1-e^{-y}\right)$

$$=1-e^{-x}-e^{-y}+e^{-\left(x+y\right)}$$

The joint probability density function is given by

$$f\left(x,y\right)=\frac{\partial^2F\left(x,y\right)}{\partial x\partial y}$$

$$=\frac{\partial^2}{\partial x\partial y}\left[1-e^{-x}-e^{-y}+e^{-\left(x+y\right)}\right]$$

$$= \frac{\partial}{\partial x}\left[ e^{-y} - e^{-(x+y)} \right]$$

$$\therefore f(x,y) = \begin{cases} e^{-(x+y)}, & x \geq 0,\ y \geq 0 \\ 0 & ,\ otherwise \end{cases}$$

(ii) The marginal probability function of $X$ is given by

$$f(x) = f_X(x)$$

$$= \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{0}^{\infty} e^{-(x+y)}\,dy$$

$$= \left[ \frac{e^{-(x+y)}}{-1} \right]_0^{\infty}$$

$$= \left[ -e^{-(x+y)} \right]_0^{\infty}$$

$$= e^{-x},\ x > 0$$

The marginal probability function of $Y$ is

$$f(y) = f_Y(y)$$

$$= \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$= \int_{0}^{\infty} e^{-(x+y)}\,dx = \left[ -e^{-(x+y)} \right]_0^{\infty}$$

$$= e^{-y},\ y > 0$$

$$\therefore f(x)f(y) = e^{-x}e^{-y} = e^{-(x+y)} = f(x,y)$$

$\therefore$ $X$ and $Y$ are independent.

(iii) $P(1 < X < 3, 1 < Y < 2) = P(1 < X < 3) \times P(1 < Y < 2)$ [Since $X$ and $Y$ are independent]

$$= \int_1^3 f(x)dx \times \int_1^2 f(y)\,dy.$$

$$= \int_1^3 e^{-x}dx \times \int_1^2 e^{-y}dy$$

$$= \left[ \frac{e^{-x}}{-1} \right]_1^3 \times \left[ \frac{e^{-y}}{-1} \right]_1^2$$

$$= \left(-e^{-3} + e^{-1}\right)\left(-e^{-2} + e^{-1}\right)$$

$$= e^{-5} - e^{-4} - e^{-3} + e^{-2}$$

b). $P\ 1\quad Y\quad {}^{3}\!/_{X}\quad 2 \qquad \int_1^3 f\ {}^{y}\!/_{x}\quad 2\ dy$

$$f_X(x) = \int_2^4 f(x,y)\,dy$$

$$= \int_2^4 \left( \frac{6 - x - y}{8} \right) dy$$

$$= \frac{1}{8} \left( 6y - xy - \frac{y^2}{2} \right)_2^4$$

$$= \frac{1}{8} (16 - 4x - 10 + 2x)$$

$$f\left( {}^{y}\!\diagup\!{}_{x} \right) = \frac{f(x, y)}{f(x)} = \frac{\dfrac{6 - x - y}{8}}{\dfrac{6 - 2x}{8}} = \frac{6 - x - y}{6 - 2x}$$

$$P \;\; 1 \quad Y \quad {}^{3}\!\diagup\!{}_{X} \quad 2 \qquad {}_{1}^{3} \; f \;\; {}^{y}\!\diagup\!{}_{x} \quad 2 \;\; dy$$

$$= \int_2^3 \left( \frac{4 - y}{2} \right) dy$$

$$= \frac{1}{2} \left[ 4y - \frac{y^2}{2} \right]_2^3$$

$$= \frac{1}{2} \left[ 4y - \frac{y^2}{2} \right]_2^3 = \frac{1}{2} \left[ 14 - \frac{17}{2} \right] = \frac{11}{4} .$$

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF SCIENCE AND HUMANITIES

DEPARTMENT OF MATHEMATICS

# UNIT – IV - Correlation and Regression – SMTA1402

## SMTA1402 - Probability and Statistics

## Unit-4 Correlation and Regression

21).a). Two random variables $X$ and $Y$ have the following joint probability density function $f(x,y)=\begin{cases}2-x-y, & 0\le x\le 1,\ 0\le y\le 1\\ 0 & , otherwise\end{cases}$. Find the marginal probability density function of $X$ and $Y$. Also find the covariance between $X$ and $Y$.

b). If $f(x,y)=\dfrac{6-x-y}{8}, 0\le x\le 2, 2\le y\le 4$ for a bivariate $X,Y$, find the correlation coefficient

Solution:

a) Given the joint probability density function $f(x,y)=\begin{cases}2-x-y, & 0\le x\le 1,\ 0\le y\le 1\\ 0 & , otherwise\end{cases}$

Marginal density function of $X$ is $f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$

$$=\int_0^1 (2-x-y)\,dy$$

$$=\left[2y-xy-\frac{y^2}{2}\right]_0^1$$

$$=2-x-\frac{1}{2}$$

$$f_X(x)=\begin{cases}\frac{3}{2}-x, & 0<x\le 1\\ 0 & , \quad otherwise\end{cases}$$

Marginal density function of $Y$ is $f_Y(y)=\int_0^1 (2-x-y)\,dx$

$$=\left[2x-\frac{x^2}{2}-xy\right]_0^1$$

$$=\frac{3}{2}-y$$

$$f_Y(y)=\begin{cases}\frac{3}{2}-y, & 0\le y\le 1\\ 0 & , \quad otherwise\end{cases}$$

Covariance of $(X,Y)=Cov(X,Y)=E(XY)-E(X)E(Y)$

$$E(X)=\int_0^1 xf_X(x)\,dx=\int_0^1 x\left(\frac{3}{2}-x\right)dx=\left[\frac{3}{2}\frac{x^2}{2}-\frac{x^3}{3}\right]_0^1=\frac{5}{12}$$

$$E(Y)=\int_0^1 yf_Y(y)\,dy=\int_0^1 y\left(\frac{3}{2}-y\right)dy=\frac{5}{12}$$

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

$$E(XY)=\int_0^1\int_0^1 xy\,f(x,y)\,dxdy$$

$$=\int_0^1\int_0^1 xy(2-x-y)\,dxdy$$

$$=\int_0^1\int_0^1 \left(2xy-x^2y-xy^2\right)dxdy$$

$$=\int_0^1\left[\frac{2x^2y}{2}-\frac{x^3}{3}y-\frac{x^2}{2}y^2\right]_0^1 dy$$

$$=\int_0^1\left(y-\frac{1}{3}-\frac{y^2}{2}\right)dy$$

$$=\left[\frac{y^2}{2}-\frac{y}{3}-\frac{y^3}{6}\right]_0^1=\frac{1}{6}$$

$$Cov(X,Y)=\frac{1}{6}-\frac{5}{12}\times\frac{5}{12}$$

$$=\frac{1}{6}-\frac{25}{144}=-\frac{1}{144}.$$

b). Correlation coefficient $\rho_{XY} = \dfrac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$

Marginal density function of $X$ is

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{2}^{4} \left(\frac{6-x-y}{8}\right) dy = \frac{6-2x}{8}$$

Marginal density function of $Y$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{0}^{2} \left(\frac{6-x-y}{8}\right) dx = \frac{10-2y}{8}$$

Then $E(X) = \int_0^2 x f_X(x)\,dx = \int_0^2 x\left(\dfrac{6-2x}{8}\right)dx$

$$= \frac{1}{8}\left[\frac{6x^2}{2} - \frac{2x^3}{3}\right]_0^2$$

$$= \frac{1}{8}\left[12 - \frac{16}{13}\right] = \frac{1}{8} \times \frac{20}{3} = \frac{5}{6}$$

$$E(Y) = \int_2^4 y\left(\frac{10-2y}{8}\right)dy = \frac{1}{8}\left[\frac{10y^2}{2} - \frac{2y^3}{3}\right]_2^4 \quad = \frac{17}{6}$$

$$E(X^2) = \int_0^2 x^2 f_x(x)\,dx \quad = \int_0^2 x^2\left(\frac{6-2x}{8}\right)dx = \frac{1}{8}\left[\frac{6x^3}{3} - \frac{2x^4}{4}\right]_0^2 = 1$$

$$E(Y^2) = \int_2^4 y^2\left(\frac{10-2y}{8}\right)dy = \frac{1}{8}\left[\frac{10y^3}{3} - \frac{2y^4}{4}\right]_2^4 = \frac{25}{3}$$

$$Var(X) = \sigma_X^2 = E(X^2) - \left[E(X)\right]^2 = 1 - \left(\frac{5}{6}\right)^2 = \frac{11}{36}$$

$$Var(Y) = \sigma_Y^2 = E(Y^2) - \left[E(Y)\right]^2 = \frac{25}{3} - \left(\frac{17}{6}\right)^2 = \frac{11}{36}$$

$$E(XY) = \int_2^4 \int_0^2 xy\left(\frac{6-x-y}{8}\right)dx\,dy$$

$$= \frac{1}{8}\int_2^4 \left[\frac{6x^2y}{2} - \frac{x^3y}{3} - \frac{x^2y^2}{2}\right]_0^2 dy$$

$$= \frac{1}{8}\int_2^4 \left(12y - \frac{8}{3}y - 2y^2\right)dy = \frac{1}{8}\left[\frac{12y^2}{2} - \frac{8}{3}\frac{y^2}{2} - \frac{2y^3}{3}\right]_2^4$$

$$= \frac{1}{8}\left[96 - \frac{64}{3} - \frac{128}{3} - 24 + \frac{16}{3} + \frac{16}{3}\right] = \frac{1}{8}\left[\frac{56}{3}\right]$$

$$E(XY) = \frac{7}{3}$$

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} = \frac{\frac{7}{3} - \left(\frac{5}{6}\right)\left(\frac{17}{6}\right)}{\frac{\sqrt{11}}{6}\frac{\sqrt{11}}{6}}$$

$$\rho_{XY} = -\frac{1}{11}.$$

22.a). Let the random variables $X$ and $Y$ have pdf $f(x,y) = \frac{1}{3}$, $(x,y) = (0,0),(1,1),(2,0)$. Compute the correlation coefficient.

b) Let $X_1$ and $X_2$ be two independent random variables with means 5 and 10 and standard devotions 2 and 3 respectively. Obtain the correlation coefficient of $UV$ where $U = 3X_1 + 4X_2$ and $V = 3X_1 - X_2$.

Solution:

a). The probability distribution is

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

$$E(X) = \sum_i x_i p_i(x_i) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(2 \times \frac{1}{3}\right) = 1$$

$$E(Y) = \sum_j y_i p_j(y_j) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(0 \times \frac{1}{3}\right) = \frac{1}{3}$$

$$E(X^2) = \sum_i x_i^2 p(x_i) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(4 \times \frac{1}{3}\right) = \frac{5}{3}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{5}{3} - 1 = \frac{2}{3}$$

$$E(Y^2) = \sum_j y_j^2 p(y_j) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right) + \left(0 \times \frac{1}{3}\right) = \frac{1}{3}$$

$$\therefore V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

Correlation coefficient $\rho_{XY} = \frac{E(XY)-E(X)E(Y)}{\sqrt{V(X)V(Y)}}$

$$E(XY) = \sum_i \sum_j x_i y_j p(x_i, y_j)$$

$$= 0.0.\frac{1}{3} + 0.1.0 + 1.0.0 + 1.1.\frac{1}{3} + 1.2.0 + 0.0.0 + 0.1.0 + 0.2.\frac{1}{3} = \frac{1}{3}$$

$$\rho_{XY} = \frac{\frac{1}{3} - (1)\left(\frac{1}{3}\right)}{\sqrt{\frac{2}{3} \times \frac{2}{9}}} = 0$$

Correlation coefficient $= 0$.

b). Given $E(X_1) = 5, \quad E(X_2) = 10$

$V(X_1) = 4, \quad V(X_2) = 9$

Since $X$ and $Y$ are independent $E(XY) = E(X)E(Y)$

Correlation coefficient $= \frac{E(UV)-E(U)E(V)}{\sqrt{Var(U)Var(V)}}$

$$E(U) = E(3X_1 + 4X_2) = 3E(X_1) + 4E(X_2)$$

$$= (3\times5) + (4\times10) = 15 + 40 = 55.$$

$$E(V) = E(3X_1 - X_2) = 3E(X_1) - E(X_2)$$

$$= (3\times5) - 10 = 15 - 10 = 5$$

$$E(UV) = E\left[(3X_1 + 4X_2)(3X_1 - X_2)\right]$$

$$= E\left[9X_1^2 - 3X_1X_2 + 12X_1X_2 - 4X_2^2\right]$$

$$= 9E(X_1^2) - 3E(X_1X_2) + 12E(X_1X_2) - 4E(X_2^2)$$

$$= 9E(X_1^2) + 9E(X_1X_2) - 4E(X_2^2)$$

$$= 9E(X_1^2) + 9E(X_1)E(X_2) - 4E(X_2^2)$$

$$= 9E(X_1^2) + 450 - 4E(X_2^2)$$

$$V(X_1) = E(X_1^2) - \left[E(X_1)\right]^2$$

$$E(X_1^2) = V(X_1) + \left[E(X_1)\right]^2 = 4 + 25 = 29$$

$$E(X_2^2) = V(X_2) + \left[E(X_2)\right]^2 = 9 + 100 = 109$$

$$\therefore E(UV) = (9\times29) + 450 - (4\times109)$$

$$= 261 + 450 - 436 = 275$$

$$Cov(U,V) = E(UV) - E(U)E(V)$$

$$= 275 - (5\times 55) = 0$$

Since $Cov(U,V)=0$, Correlation coefficient $=0$.

23.a). Let the random variable $X$ has the marginal density function $f(x)=1,\ -\frac{1}{2}<x<\frac{1}{2}$ and let the conditional density of $Y$ be $f\left(\frac{y}{x}\right)=\begin{cases}1, & x<y<x+1,\ -\frac{1}{2}<x<0\\ 1, & -x<y<1-x,\ 0<x<\frac{1}{2}\end{cases}$. Prove that the variables $X$ and $Y$ are uncorrelated.

b). Given $f(x,y)=xe^{-x(y+1)},\ x\geq 0,\ y\geq 0$. Find the regression curve of $Y$ on $X$.

Solution:

a). We have $E(X)=\int_{-\frac{1}{2}}^{\frac{1}{2}} xf(x)dx=\int_{-\frac{1}{2}}^{\frac{1}{2}} xdx=\left[\frac{x^2}{2}\right]_{-\frac{1}{2}}^{\frac{1}{2}}=0$

$$E(XY)=\int_{-\frac{1}{2}}^{0}\int_{x}^{x+1} xydxdy+\int_{0}^{\frac{1}{2}}\int_{-x}^{1-x} xydxdy$$

$$=\int_{-\frac{1}{2}}^{0} x\left[\int_{x}^{x+1} ydy\right]dx+\int_{0}^{\frac{1}{2}} x\left[\int_{-x}^{1-x} ydy\right]dx$$

$$=\frac{1}{2}\int_{-\frac{1}{2}}^{0} x(2x+1)dx+\frac{1}{2}\int_{0}^{\frac{1}{2}} x(1-2x)dx$$

$$=\frac{1}{2}\left[\frac{2x^3}{3}+\frac{x^2}{2}\right]_{-\frac{1}{2}}^{0}+\frac{1}{2}\left[\frac{x^2}{2}-\frac{2x^3}{3}\right]_{0}^{\frac{1}{2}}=0$$

Since $Cov(X,Y)=E(XY)-E(X)E(Y)=0$, the variables $X$ and $Y$ are uncorrelated.

b). Regression curve of $Y$ on $X$ is $E\left(\frac{y}{x}\right)$

$$E\left(\frac{y}{x}\right)=\int_{-\infty}^{\infty} yf\left(\frac{y}{x}\right)dy$$

$$f(y/x)=\frac{f(x,y)}{f_X(X)}$$

Marginal density function $f_X(x)=\int_{0}^{\infty} f(x,y)dy$

$$= x\int_0^\infty e^{-x(y+1)}dy$$

$$= x\left[\frac{e^{-x(y+1)}}{-x}\right]_0^\infty = e^{-x}, \quad x \ge 0$$

Conditional pdf of $Y$ on $X$ is $f\left(y/x\right) = \frac{f(x,y)}{f_X(x)} = \frac{xe^{-xy-x}}{e^{-x}} = xe^{-xy}$

The regression curve of $Y$ on $X$ is given by

$$E\left(y/x\right) = \int_0^\infty yxe^{-xy}dy$$

$$= x\left[y\frac{e^{-xy}}{-x} - \frac{e^{-xy}}{x^2}\right]_0^\infty$$

$E\left(y/x\right) = \frac{1}{x} \Rightarrow y = \frac{1}{x}$ and hence $xy = 1$.

24.a). Given $f(x,y) = \begin{cases} \frac{x+y}{3}, & 0 < x < 1,\ 0 < y < 2 \\ 0 & , otherwise \end{cases}$, obtain the regression of $Y$ on $X$ and $X$ on $Y$.

b). Distinguish between correlation and regression Analysis

Solution:

a). Regression of Y on X is $E\left(Y/X\right)$

$$E\left(Y/X\right) = \int_{-\alpha}^{\alpha} yf\left(y/x\right)dy$$

$$f\left(Y/X\right) = \frac{f(x,y)}{f_X(x)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$= \int_0^2 \left(\frac{x+y}{3}\right)dy = \frac{1}{3}\left[xy + \frac{y^2}{2}\right]_0^2$$

$$= \frac{2(x+1)}{3}$$

$$f\left(Y/X\right) = \frac{f(x,y)}{f_X(x)} = \frac{x+y}{2(x+1)}$$

Regression of $Y$ on $X$ = $E\left(Y/X\right) = \int_0^2 \frac{y(x+y)}{2(x+1)}dy$

$$= \frac{1}{2(x+1)}\left[\frac{xy^2}{2}+\frac{y^3}{3}\right]_0^2$$

$$= \frac{1}{2(x+1)}\left[2x+\frac{8}{3}\right] = \frac{3x+4}{3(x+1)}$$

$$E\left(X/Y\right) = \int_{-\infty}^{\infty} xf\left(x/y\right)dx$$

$$f\left(x/y\right) = \frac{f(x,y)}{f_Y(y)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$= \int_0^1 \left(\frac{x+y}{3}\right)dx = \frac{1}{3}\left[\frac{x^2}{2}+xy\right]_0^1$$

$$= \frac{1}{3}\left[\frac{1}{2}+y\right]$$

$$f\left(x/y\right) = \frac{2(x+y)}{2y+1}$$

Regression of $X$ on $Y = E\left(X/Y\right) = \int_0^1 \frac{x+y}{2y+1}dx$

$$= \frac{1}{2y+1}\left[\frac{x^2}{2}+xy\right]_0^1$$

$$= \frac{\frac{1}{2}+y}{2y+1} = \frac{1}{2}.$$

b).

1. Correlation means relationship between two variables and Regression is a Mathematical Measure of expressing the average relationship between the two variables.
2. Correlation need not imply cause and effect relationship between the variables. Regression analysis clearly indicates the cause and effect relationship between Variables.
3. Correlation coefficient is symmetric i.e. $r_{xy} = r_{yx}$ where regression coefficient is not symmetric
4. Correlation coefficient is the measure of the direction and degree of linear relationship between two variables. In regression using the relationship between two variables we can predict the dependent variable value for any given independent variable value.

25.a). $X$ any $Y$ are two random variables with variances $\sigma_x^2$ and $\sigma_y^2$ respectively and $r$ is the coefficient of correlation between them. If $U = X + KY$ and $V = X + \frac{y\sigma_x}{\sigma_y}$, find the value of $k$ so that $U$ and $V$ are uncorrelated.

b). Find the regression lines:

| $X$ | 6 | 8 | 10 | 18 | 20 | 23 |
|---|---|---|---|---|---|---|
| $Y$ | 40 | 36 | 20 | 14 | 10 | 2 |

Solution:

Given $U = X + KY$

$$E(U) = E(X) + KE(Y)$$

$$V = X + \frac{\sigma_X}{\sigma_Y} Y$$

$$E(V) = E(X) + \frac{\sigma_X}{\sigma_Y} E(Y)$$

If $U$ and $V$ are uncorrelated, $Cov(U,V) = 0$

$$E\left[\left(U - E(U)\right)\left(V - E(V)\right)\right] = 0$$

$$\Rightarrow E\left[\left(X + KY - E(X) - KE(Y)\right) \times \left(X + \frac{\sigma_X}{\sigma_Y} Y - E(X) - \frac{\sigma_X}{\sigma_Y} E(Y)\right)\right] = 0$$

$$\Rightarrow E\left\{\left[\left(X - E(X)\right) + K\left(Y - E(Y)\right)\right] \times \left[\left(X - E(X)\right) + \frac{\sigma_X}{\sigma_Y}\left(Y - E(Y)\right)\right]\right\} = 0$$

$$\Rightarrow E\left\{\left(X - E(X)\right)^2 + \frac{\sigma_X}{\sigma_Y}\left(X - E(X)\right)\left(Y - E(Y)\right) + K\left(Y - E(Y)\right)\left(X - E(X)\right) + K\frac{\sigma_X}{\sigma_Y}\left(Y - E(Y)\right)^2\right\} = 0$$

$$\Rightarrow V(X) + \frac{\sigma_X}{\sigma_Y} Cov(X,Y) + KCov(X,Y) + K\frac{\sigma_X}{\sigma_Y} V(Y) = 0$$

$$K\left[Cov(X,Y) + \frac{\sigma_X}{\sigma_Y} V(Y)\right] = -V(X) - \frac{\sigma_X}{\sigma_Y} Cov(x,y)$$

$$K = \frac{-V(X) - \frac{\sigma_X}{\sigma_Y} r\sigma_X\sigma_Y}{r\sigma_X\sigma_Y + \frac{\sigma_X}{\sigma_Y} V(Y)} = \frac{-\sigma_X^2 - r\sigma_X^2}{r\sigma_X\sigma_Y + \sigma_X\sigma_Y}$$

$$= \frac{-\sigma_X^2(1+r)}{\sigma_X\sigma_Y(1+r)} = -\frac{\sigma_X}{\sigma_Y}.$$

b).

| $X$ | $Y$ | $X^2$ | $Y^2$ | $XY$ |
|---|---|---|---|---|
| 6 | 40 | 36 | 1600 | 240 |

| | | | | |
|---|---|---|---|---|
| 8 | 36 | 64 | 1296 | 288 |
| 10 | 20 | 100 | 400 | 200 |
| 18 | 14 | 324 | 196 | 252 |
| 20 | 10 | 400 | 100 | 200 |
| 23 | 2 | 529 | 4 | 46 |
| $\sum X = 85$ | $\sum Y = 122$ | $\sum X^2 = 1453$ | $\sum Y^2 = 3596$ | $\sum XY = 1226$ |

$$\overline{X} = \frac{\sum x}{n} = \frac{85}{6} = 14.17,\ \overline{Y} = \frac{\sum y}{n} = \frac{122}{6} = 20.33$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{1453}{6} - \left(\frac{85}{6}\right)^2} = 6.44$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{3596}{6} - \left(\frac{122}{6}\right)^2} = 13.63$$

$$r = \frac{\frac{\sum xy}{n} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} = \frac{\frac{1226}{6} - (14.17)(20.33)}{(6.44)(13.63)} = -0.95$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = -0.95 \times \frac{6.44}{13.63} = -0.45$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = -0.95 \times \frac{13.63}{6.44} = -2.01$$

The regression line $X$ on $Y$ is

$$x - \overline{x} = b_{xy}\left(y - \overline{y}\right) \Rightarrow x - 14.17 = -0.45\left(y - \overline{y}\right)$$

$$\Rightarrow x = -0.45y + 23.32$$

The regression line $Y$ on $X$ is

$$y - \overline{y} = b_{yx}\left(x - \overline{x}\right) \Rightarrow y - 20.33 = -2.01(x - 14.17)$$

$$\Rightarrow y = -2.01x + 48.81$$

26. a) Using the given information given below compute $\overline{x}, \overline{y}$ and $r$. Also compute $\sigma_y$ when $\sigma_x = 2$, $2x + 3y = 8$ and $4x + y = 10$.

b) The joint pdf of $X$ and $Y$ is

| $Y$ \ $X$ | -1 | 1 |
|---|---|---|

| | | |
|---|---|---|
| 0 | $\frac{1}{8}$ | $\frac{3}{8}$ |
| 1 | $\frac{2}{8}$ | $\frac{2}{8}$ |

Find the correlation coefficient of $X$ and $Y$.

Solution:

a). When the regression equation are Known the arithmetic means are computed by solving the equation.

$$2x+3y=8 \text{------------ (1)}$$

$$4x+y=10 \text{------------ (2)}$$

$$(1)\times 2 \Rightarrow 4x+6y=16 \text{------- (3)}$$

$$(2)-(3) \Rightarrow -5y=-6$$

$$\Rightarrow y=\frac{6}{5}$$

Equation $(1) \Rightarrow 2x+3\left(\frac{6}{5}\right)=8$

$$\Rightarrow 2x=8-\frac{18}{5}$$

$$\Rightarrow x=\frac{11}{5}$$

i.e. $\bar{x}=\frac{11}{5}$ & $\bar{y}=\frac{6}{5}$

To find $r$, Let $2x+3y=8$ be the regression equation of $X$ on $Y$.

$$2x=8-3y \Rightarrow x=4-\frac{3}{2}y$$

$\Rightarrow b_{xy}=$Coefficient of $Y$ in the equation of $X$ on $Y=-\frac{3}{2}$

Let $4x+y=10$ be the regression equation of $Y$ on $X$

$\Rightarrow y=10-4x$

$\Rightarrow b_{yx}=$coefficient of $X$ in the equation of $Y$ on $X=-4$.

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

$$=-\sqrt{\left(-\frac{3}{2}\right)(-4)} \qquad (\because b_{xy} \text{ \& } b_{yx} \text{ are negative })$$

$$=-2.45$$

Since $r$ is not in the range of $(-1\le r\le 1)$ the assumption is wrong.

Now let equation $(1)$ be the equation of $Y$ on $X$

$$\Rightarrow y=\frac{8}{3}-\frac{2x}{3}$$

$\Rightarrow b_{yx}=$Coefficient of $X$ in the equation of $Y$ on $X$

$b_{yx} = -\frac{2}{3}$

from equation (2) be the equation of $X$ on $Y$

$b_{xy} = -\frac{1}{4}$

$r = \pm\sqrt{b_{xy} b_{yx}} \qquad = \sqrt{-\frac{2}{3} \times -\frac{1}{4}} = 0.4081$

To compute $\sigma_y$ from equation $(4)$ $b_{yx} = -\frac{2}{3}$

But we know that $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$$\Rightarrow -\frac{2}{3} = 0.4081 \times \frac{\sigma_y}{2}$$

$$\Rightarrow \sigma_y = -3.26$$

b). Marginal probability mass function of $X$ is

When $X = 0, \quad P(X) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$

$X = 1, \quad P(X) = \frac{2}{8} + \frac{2}{8} = \frac{4}{8}$

Marginal probability mass function of $Y$ is

When $Y = -1, \quad P(Y) = \frac{1}{8} + \frac{2}{8} = \frac{3}{8}$

$Y = 1, \quad P(Y) = \frac{3}{8} + \frac{2}{8} = \frac{5}{8}$

$$E(X) = \sum_x x\, p(x) = 0 \times \frac{4}{8} + 1 \times \frac{4}{8} = \frac{4}{8}$$

$$E(Y) = \sum_y y\, p(y) = -1 \times \frac{3}{8} + 1 \times \frac{5}{8} = -\frac{3}{8} + \frac{5}{8} = \frac{2}{8}$$

$$E(X^2) = \sum_x x^2\, p(x) = 0^2 \times \frac{4}{8} + 1^2 \times \frac{4}{8} = \frac{4}{8}$$

$$E(Y^2) = \sum_y y^2\, p(y) = (-1)^2 \times \frac{3}{8} + 1^2 \times \frac{5}{8} = \frac{3}{8} + \frac{5}{8} = 1$$

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

$$= \frac{4}{8} - \left(\frac{4}{8}\right)^2 = \frac{1}{4}$$

$$V(Y) = E(Y^2) - (E(Y))^2$$

$$= 1 - \left(\frac{1}{4}\right)^2 = \frac{15}{16}$$

$$E(XY)=\sum_x\sum_y xy\,p(x,y)$$

$$=0\times\frac{1}{8}+0\times\frac{3}{8}+(-1)\frac{2}{8}+1\times\left(\frac{2}{8}\right)=0$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0-\frac{1}{2}\times\frac{1}{4}=-\frac{1}{8}$$

$$r=\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}=\frac{-\frac{1}{8}}{\sqrt{\frac{1}{4}}\sqrt{\frac{15}{16}}}=-0.26\,.$$

27. a) Calculate the correlation coefficient for the following heights (in inches) of fathers $X$ and their sons $Y$.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $X$ | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 |
| $Y$ | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |

b) If $X$ and $Y$ are independent exponential variates with parameters 1, find the pdf of $U=X-Y$.

Solution:

| $X$ | $Y$ | $XY$ | $X^2$ | $Y^2$ |
|---|---|---|---|---|
| 65 | 67 | 4355 | 4225 | 4489 |
| 66 | 68 | 4488 | 4359 | 4624 |
| 67 | 65 | 4355 | 4489 | 4285 |
| 68 | 72 | 4896 | 4624 | 5184 |
| 69 | 72 | 4968 | 4761 | 5184 |
| 70 | 69 | 4830 | 4900 | 4761 |
| 72 | 71 | 5112 | 5184 | 5041 |
| $\sum X=544$ | $\sum Y=552$ | $\sum XY=37560$ | $\sum X^2=37028$ | $\sum Y^2=38132$ |

$$\overline{X}=\frac{\sum x}{n}=\frac{544}{8}=68$$

$$\overline{Y}=\frac{\sum y}{n}=\frac{552}{8}=69$$

$$\overline{X}\,\overline{Y}=68\times69=4692$$

$$\sigma_X=\sqrt{\frac{1}{n}\sum x^2-\overline{X}^2}=\sqrt{\frac{1}{8}(37028)-68^2}=\sqrt{4628.5-4624}=2.121$$

$$\sigma_Y=\sqrt{\frac{1}{n}\sum y^2-y^2}=\sqrt{\frac{1}{8}(38132)-69^2}=\sqrt{4766.5-4761}=2.345$$

$$Cov(X,Y)=\frac{1}{n}\sum XY-\overline{X}\,\overline{Y}=\frac{1}{8}(37650)-68\times 69$$
$$=4695-4692=3$$

The correlation coefficient of $X$ and $Y$ is given by

$$r(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{3}{(2.121)(2.345)}$$
$$=\frac{3}{4.973}=0.6032\,.$$

b). Given that $X$ and $Y$ are exponential variates with parameters 1

$$f_X(x)=e^{-x},\ x\ge 0,\ f_Y(y)=e^{-y},\ y\ge 0$$

Also $f_{XY}(x,y)=f_X(x)f_y(y)$ since $X$ and $Y$ are independent

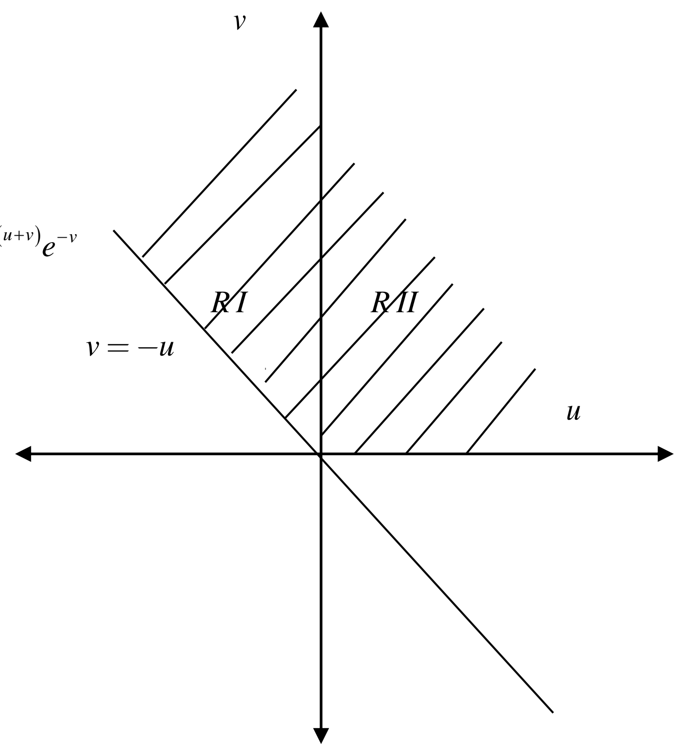
$$=e^{-x}e^{-y}$$
$$=e^{-(x+y)};\ x\ge 0,\ y\ge 0$$

Consider the transformations $u=x-y$ and $v=y$

$\Rightarrow x=u+v\,,\ y=v$

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}1 & -1\\ 0 & 1\end{vmatrix}=1$$

$$f_{UV}(u,v)=f_{XY}(x,y)\left|J\right|=e^{-x}e^{-y}=e^{-(u+v)}e^{-v}$$
$$=e^{-(u+2v)},\ u+v\ge 0,\ v\ge 0$$




In Region I when $u<0$

$$f(u)=\int_{-u}^{\infty} f(u,v)\,dv=\int_{-u}^{\infty} e^{-u}.e^{-2v}dv$$
$$=e^{-u}\left[\frac{e^{-2v}}{-2}\right]_{-u}^{\infty}$$
$$=\frac{e^{-u}}{-2}\left[0-e^{2u}\right]=\frac{e^{u}}{2}$$

In Region II when $u>0$

$$f(u)=\int_{0}^{\infty} f(u,v)dv$$
$$=\int_{0}^{\infty} e^{-(u+2v)}dv=\frac{e^{-u}}{2}$$

$$\therefore f(u)=\begin{cases}\frac{e^{u}}{2}, & u<0\\ \frac{e^{-u}}{2}, & u>0\end{cases}$$

28. a) The joint pdf of $X$ and $Y$ is given by $f(x,y)=e^{-(x+y)}, x>0, y>0$. Find the pdf of $U=\dfrac{X+Y}{2}$.

b) If $X$ and $Y$ are independent random variables each following $N(0,2)$, find the pdf of $Z=2X+3Y$. If $X$ and $Y$ are independent rectangular variates on $(0,1)$ find the distribution of $\dfrac{X}{Y}$.

Solution:

a). Consider the transformation $u=\dfrac{x+y}{2}$ & $v=y$

$$\Rightarrow x=2u-v \text{ and } y=v$$

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}2 & -1\\ 0 & 1\end{vmatrix}=2$$

$$f_{UV}(u,v)=f_{XY}(x,y)|J|$$

$$=e^{-(x+y)}2=2e^{-(x+y)}=2e^{-(2u-v+v)}$$

$$=2e^{-2u},\ 2u-v\geq 0,\ v\geq 0$$

$$f_{UV}(u,v)=2e^{-2u},\ u\geq 0,\ 0\leq v\leq \frac{u}{2}$$

$$f(u)=\int_0^{\frac{u}{2}} f_{UV}(u,v)\,dv=\int_0^{\frac{u}{2}} 2e^{-2u}dv$$

$$=\left[2e^{-2u}v\right]_0^{\frac{u}{2}}$$

$$f(u)=\begin{cases}2\frac{u}{2}e^{-2u}, & u\geq 0\\ 0 & , otherwise\end{cases}$$

b).(i) Consider the transformations $w=y$,

i.e. $z=2x+3y$ and $w=y$

i.e. $x=\dfrac{1}{2}(z-3w)$, $y=w$

$$|J|=\frac{\partial(x,y)}{\partial(z,w)}=\begin{vmatrix}\frac{\partial x}{\partial z} & \frac{\partial x}{\partial w}\\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w}\end{vmatrix}=\begin{vmatrix}\frac{1}{2} & -\frac{3}{2}\\ 0 & 1\end{vmatrix}=\frac{1}{2}.$$

Given that $X$ and $Y$ are independent random variables following $N(0,2)$

$$\therefore f_{XY}(x,y)=\frac{1}{8\pi}e^{\frac{-\left(x^2+y^2\right)}{8}},-\infty<x,y<\infty$$

The joint pdf of $(z,w)$ is given by

$$f_{ZW}(z,w)=|J|\,f_{XY}(x,y)$$

$$=\frac{1}{2}\cdot\frac{1}{8\pi}e^{\frac{-\left[\frac{1}{4}(z-3w)^2+w^2\right]}{8}}$$

$$=\frac{1}{16\pi}e^{-\frac{1}{32}\left[(z-3w)^2+4w^2\right]},-\infty<z,w<\infty\,.$$

The pdf of $z$ is the marginal pdf obtained by interchanging $f_{ZW}(z,w)$ w.r.to $w$ over the range of $w$.

$$\therefore f_Z(z)=\frac{1}{16\pi}\int_{-\infty}^{\infty}\left(e^{-\frac{1}{32}\left(z^2-6wz+13w^2\right)}\right)dw$$

$$=\frac{1}{16\pi}e^{-\frac{z^2}{32}}\int_{-\infty}^{\infty}\left(e^{-\frac{13}{32}\left(w^2-\frac{6wz}{13}+\left(\frac{3z}{13}\right)^2-\left(\frac{3z}{13}\right)^2\right)}\right)dw$$

$$=\frac{1}{16\pi}e^{-\frac{z^2}{32}+\frac{9z^2}{13\times32}}\int_{-\infty}^{\infty}\left(e^{-\frac{13}{32}\left(w-\frac{3z}{13}\right)^2}\right)dw$$

$$=\frac{1}{16\pi}e^{-\frac{z^2}{8\times13}}\int_{-\infty}^{\infty}e^{-\frac{13}{32}t^2}dt$$

$$r=\frac{13}{32}t^2\Rightarrow dr=\frac{13}{16}tdt\Rightarrow\frac{16}{13t}dr=dt\Rightarrow\sqrt{\frac{r32}{13}}dr=dt$$

$$\frac{16}{13}\sqrt{\frac{13}{r32}}dr=dt\Rightarrow\frac{4}{\sqrt{13}\times\sqrt{2}}r^{-\frac{1}{2}}dr=dt$$

$$=\frac{2}{16\pi}\frac{4}{\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times13}}\int_{0}^{\infty}e^{-r}r^{-\frac{1}{2}}dr$$

$$=\frac{1}{2\pi\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times13}}\int_{0}^{\infty}e^{-r}r^{-\frac{1}{2}}dr$$

$$=\frac{1}{2\pi\sqrt{13}\times\sqrt{2}}e^{-\frac{z^2}{8\times13}}\sqrt{\pi}=\frac{1}{2\sqrt{13}\sqrt{2\pi}}e^{-\frac{z^2}{2\left(2\sqrt{13}\right)^2}}$$

i.e. $Z\sim N\left(0,2\sqrt{13}\right)$

b).(ii) Given that $X$ and $Y$ are uniform Variants over $(0,1)$

$$\therefore f_X(x)=\begin{cases}1, & 0<x<1\\ 0, & otherwise\end{cases}\text{ and }f_Y(y)=\begin{cases}1, & 0<y<1\\ 0, & otherwise\end{cases}$$

Since $X$ and $Y$ are independent,

$$f_{XY}(x,y) = f_X(x) f_y(y) \begin{cases} 1, & 0 < x, y < 1 \\ 0, & otherwise \end{cases}$$

Consider the transformation $u = \dfrac{x}{y}$ and $v = y$

i.e. $x = uv$ and $y = v$

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix} = v$$

$$\therefore f_{UV}(u,v) = f_{XY}(x,y)|J|$$
$$= v,\ 0 < u < \infty,\ 0 < v < \infty$$

The range for $u$ and $v$ are identified as follows.

$0 < x < 1$ and $0 < y < 1$.

$\Rightarrow 0 < uv < 1$ and $0 < v < 1$

$\Rightarrow uv > 0,\ uv < 1,\ v > 0$ and $v < 1$

$\Rightarrow uv > 0$ and $v > 0 \Rightarrow u > 0$

Now $f(u) = \int f_{UV}(u,v)\,dv$

The range for $v$ differs in two regions

$$f(u) = \int_0^1 f_{UV}(u,v)\,dv$$

$$= \int_0^1 v\,dv = \left[\frac{v^2}{2}\right]_0^1 = \frac{1}{2},\ 0 < u < 1$$

$$f(u) = \int_0^{\frac{1}{u}} f_{UV}(u,v)\,dv$$

$$= \int_0^{\frac{1}{u}} v\,dv = \left[\frac{v^2}{2}\right]_0^{\frac{1}{u}} = \frac{1}{2u^2},\ 1 \le u \le \infty$$

$$\therefore f(u) = \begin{cases} \dfrac{1}{2}, & 0 \le u \le 1 \\ \dfrac{1}{2u^2}, & u > 1 \end{cases}$$

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

**SCHOOL OF SCIENCE AND HUMANITIES**

**DEPARTMENT OF MATHEMATICS**

# UNIT – V - Analysis of Variance and Statistical Quality Control – SMTA1402

**SMTA1402 - Probability and Statistics**
**Unit - V Analysis of Variance and Statistical Quality Control**

**ANOVA (Analysis of Variance) :**

Analysis of Variance is a technique that will enable us to test for the significance of the difference among more than two sample means.

**Assumptions of analysis of variance:**

(i) The sample observations are independent
(ii) The environmental effects are additive in nature
(iii) The samples have been randomly selected from the population.
(iv) Parent population from which observations are taken in normal.

**One Way Classification (or) Completely randomized Design (C.R.D)**

The C.R.D is the simplest of all the designs, based on principles of randomization and replication. In this design, treatments are allocated at random to the experimental units over the entire experimental materials.

**Advantages of completely randomized block design:**

The advantages of completely randomized experimental design as follows:
(i) Easy to lay out. (ii) Allow flexibility (iii) Simple statistical analysis
(iv) lots of information due to missing data is smaller than with any other design

**Two Way Classification (or) Randomized Block Design (R.B.D):**

The entire experiment influences on only two factors is two way Classification.

**The basic principles of design of experiments:**

(i) Randomization (ii) Replication (iii) Local Control

**<u>Working Procedure ( One – Way classification )</u>**

**Null Hypothesis** $H_0$: There is no significance difference between the treatments.

**Alternate Hypothesis** $H_1$: There is a significance difference between the treatments.

**Analysis:**

**Step 1**: Find $N$= number of observations

**Setp 2**: Find $T$ = The total value of observations

**Step 3**: Find the correction Factor = $C.F = \frac{T^2}{N}$

**Step 4**: Calculate the total sum of squares = $\text{TSS} = \left(\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + ...\right) - C.\text{F}$

**Step 4**: Find Total Sum of Square $\text{TSS} = \left(\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + ...\right) - C.\text{F}$

**Step 5**: Column Sum of Square $SSC=\left(\frac{\left(\sum X_1\right)^2}{N_1}+\frac{\left(\sum X_2\right)^2}{N_2}+\frac{\left(\sum X_3\right)^2}{N_3}+...\right)-C.F$

Where $N_i=$ Total number of observation in each column ( $i=1,2,3,...$ )

**Step 6**: Prepare the ANOVA TABLE to calculate F-ratio.

| Source of Variation | Sum of Degrees | Degree of freedom | Mean Square | F- Ratio |
|---|---|---|---|---|
| Between Columns | SSC | c-1 | $MSC=\frac{SSC}{c-1}$ | $F_C=\frac{MSC}{MSE}$ if $MSC>MSE$ (or) |
| Error | SSE | N-c | $MSE=\frac{SSE}{N-c}$ | $F_C=\frac{MSE}{MSC}$ if $MSE>MSC$ |
| Total | | | | |

**Step 7:** Find the table value (use $\chi^2$ table)

**Step 8:** Conclusion:

Calculated value < Table Value, the we accept Null Hypothesis $H_0$ **(or)**

Calculated value > Table Value, the we rejectNull Hypothesis $H_0$

**<u>Working Procedure ( Two – Way classification )</u>**

**Null Hypothesis** $H_0$: There is no significance difference between the treatments.

**Alternate Hypothesis** $H_1$: There is a significance difference between the treatments.

**Analysis:**

**Step 1**: Find $N$= number of observations

**Setp 2**: Find $T=$ The total value of observations

**Step 3**: Find the correction Factor = $C.F=\frac{T^2}{N}$

**Step 4**: Calculate the total sum of squares = $TSS=\left(\sum X_1^2+\sum X_2^2+\sum X_3^2+...\right)-C.F$

**Step 4**: Find Total Sum of Square $TSS=\left(\sum X_1^2+\sum X_2^2+\sum X_3^2+...\right)-C.F$

**Step 5**: Find column sum of Square $SSC=\left(\frac{\left(\sum X_1\right)^2}{N_1}+\frac{\left(\sum X_2\right)^2}{N_2}+\frac{\left(\sum X_3\right)^2}{N_3}+...\right)-C.F$

Where $N_i=$ Total number of observation in each column $(i=1,2,3,...)$

**Step 6**: Find Row sum of square = $SSR=\left(\frac{\left(\sum Y_1\right)^2}{N_1}+\frac{\left(\sum Y_2\right)^2}{N_2}+\frac{\left(\sum Y_3\right)^2}{N_3}+...\right)-C.F$

Where $N_j$ = Total number of observation in each Row $(j=1,2,3,...)$

**Step 7:** Prepare the ANOVA TABLE to calculate F-ratio.

| Source of Variation | Sum of Degrees | Degree of freedom | Mean Square | F- Ratio |
|---|---|---|---|---|
| Between Columns | SSC | c-1 | $MSC=\frac{SSC}{c-1}$ | $F_C=\frac{MSC}{MSE}$ if $MSC>MSE$ (or) $F_C=\frac{MSE}{MSC}$ if $MSE>MSC$ |
| Between Rows | SSR | r-1 | $MSC=\frac{SSR}{r-1}$ | $F_R=\frac{MSR}{MSE}$ if $MSR>MSE$ (or) $F_R=\frac{MSE}{MSR}$ if $MSE>MSR$ |
| Error | SSE | N-c-r+1 | $MSE=\frac{SSE}{N-c-r+1}$ | |
| Total | TSS | rc-1 | | |

**Step 8:** Find the table value for both $F_C$ & $F_R$ (use $\chi^2$ table)

**Step 9:Conclusion:**

Calculated value < Table Value, the we accept Null Hypothesis $H_0$ **(or)**

Calculated value > Table Value, the we reject Null Hypothesis $H_0$

**1. The following are the numbers of mistakes made in 5 successive days of 4 technicians working for a photographic laboratory :**

| **Tech I ($X_1$)** | **Tech II ($X_2$)** | **Tech III ($X_3$)** | **Tech IV($X_4$)** |
|---|---|---|---|
| **6** | **14** | **10** | **9** |
| **14** | **9** | **12** | **12** |
| **10** | **12** | **7** | **8** |
| **8** | **10** | **15** | **10** |
| **11** | **14** | **11** | **11** |

**Test at the level of significance $\alpha$ = 0.01 whether the differences among the 4 samples means can be attributed to chance.**

**Solution:**

$H_0$: There is no significant difference between the technicians

$H_1$ : Significant difference between the technicians

We shift the origin

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | TOTAL | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|---|
| | -4 | 4 | 0 | -1 | **-1** | 16 | 16 | 0 | 1 |
| | 4 | -1 | 2 | 2 | **7** | 16 | 1 | 4 | 4 |
| | 0 | 2 | -3 | -2 | **-3** | 0 | 4 | 9 | 4 |
| | -2 | 0 | 5 | 0 | **3** | 4 | 0 | 25 | 0 |
| | 1 | 4 | 1 | 1 | **7** | 1 | 16 | 1 | 1 |
| Total | **-1** | **9** | **5** | **0** | **13** | **37** | **37** | **39** | **10** |

N= Total No of Observations = 20 T=Grand Total = 13

$$\text{Correction Factor} = \frac{(\text{Grand total})^2}{\text{Total No of Observations}} = 8.45$$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 37 + 37 + 39 + 10 - 8.45 = 114.55$$

$$SSC = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - C.F = \frac{(-1)^2}{5} + \frac{(9)^2}{5} + \frac{(5)^2}{5} + 0 - 8.45 = 12.95$$

SSE = TSS – SSC = 114.55-12.95= 101.6

**ANOVA Table**

| Source of Variation | Sum of Squares | Degree of freedom | Mean Square | F- Ratio |
|---|---|---|---|---|
| Between Samples | SSC=12.95 | C-1= 4-1=3 | $MSC = \frac{SSC}{K-1}$ =4.317 | $F_C = \frac{MSC}{MSE}$ =1.471 |
| Within Samples | SSE=101.6 | N-C=20-4=16 | $MSE = \frac{SSE}{N-K}$ =6.35 | |

Cal $F_C$ = 1.471 & Tab $F_C$ (16,3) = 5.29

**Conclusion :** Cal $F_C$ < Tab $F_C$ $\Rightarrow$ There is no significance difference between the technicians

2. **A completely randomized design exprement with 10 plots and 3 treatments gave the following results.**

| **Plot No** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Treatment** | **A** | **B** | **C** | **A** | **C** | **C** | **A** | **B** | **A** | **B** |
| **Yield** | **5** | **4** | **3** | **7** | **5** | **1** | **3** | **4** | **1** | **7** |

**Analyse the results for treatment effects.**

**Solution:**

| A | B | C |
|---|---|---|
| **5** | **4** | **3** |
| **7** | **4** | **5** |
| **3** | **7** | **1** |
| **1** | | |

**Null Hypothesis $H_0$:** There is no significant difference in treatments

**Alternate Hypothesis $H_1$ :** Significant difference in treatments

| | $X_1$ | $X_2$ | $X_3$ | TOTAL | $X_1^2$ | $X_2^2$ | $X_3^2$ |
|---|---|---|---|---|---|---|---|
| | 5 | 4 | 3 | **12** | 25 | 16 | 9 |
| | 7 | 4 | 5 | **16** | 49 | 16 | 25 |
| Total | 3 | 7 | 1 | **11** | 9 | 49 | 1 |
| | 1 | | | **1** | 1 | | |
| | 16 | 15 | 9 | **40** | 84 | 81 | 35 |

Step1: N= Total No of Observations = 10

Step 2: T=Grand Total = 40

Step 3: Correction Factor = $\dfrac{(\text{Grand total})^2}{\text{Total No of Observations}} = \dfrac{T^2}{N} = \dfrac{40^2}{10} = 160$

Step 4: TSS= $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 - C.\text{F} = 84 + 81 + 35 - 160 = 40$

Step 5: SSC $= \dfrac{\left(\sum X_1\right)^2}{N_1} + \dfrac{\left(\sum X_2\right)^2}{N_1} + \dfrac{\left(\sum X_3\right)^2}{N_1} - C.\text{F} = \dfrac{(16)^2}{4} + \dfrac{15^2}{3} + 3 - 160$

$SSC = 64 + 75 + 27 - 160 = 6$

Where $N_1$= Number of elements in each column

Step 7: SSE=TSS-SSC $= 40 - 6 = 34$

Step 8: **ANOVA TABLE:**

| **Source of Variation** | **Sum of Squares** | **Degree of freedom** | **Mean Square** | **F- Ratio** |
|---|---|---|---|---|
| Between Columns | SSC=6 | C-1= 3-1=2 | $\text{MSC} = \dfrac{\text{SSC}}{\text{C}-1} = \dfrac{6}{2} = 3$ | $\text{F}_\text{C} = \dfrac{\text{MSE}}{\text{MSC}} = \dfrac{4.86}{3} = 1.62$ |
| Error | SSE=34 | N-C=10-3=7 | $\text{MSE} = \dfrac{\text{SSE}}{\text{N}-\text{C}} = \dfrac{34}{87} = 4.86$ | |

Cal $\text{F}_\text{C}$ = 1.62

**Table value :** $F_C$ (7,2)=19.35

**Conclusion :** Cal $F_C$< Tab $F_C$

We accept Null Hypothesis $\Rightarrow$ There is no significance difference in tretments

**3. The following table gives the number of articles of a product produced by five different workers using four types of machines.**

| **Workers** | **Machines** | | | |
|---|---|---|---|---|
| | **P** | **Q** | **R** | **S** |
| **A** | **44** | **38** | **47** | **36** |
| **B** | **46** | **40** | **52** | **43** |
| **C** | **34** | **36** | **44** | **32** |
| **D** | **43** | **38** | **46** | **33** |
| **E** | **38** | **42** | **49** | **39** |

**Test (i) Whether the five workers differ with respect to mean productivity and**
**(ii) Whether the four machines differ with respect to mean productivity.**

**Solution: $H_0$: There is no significant difference between the Machine types and no significant difference between the Workers**

**$H_1$ :Significant difference between the Machine types and no significant difference between the Workers**

We shift the origin $X_{ij} = x_{ij} - 46$; h = 5; k = 4; N = 20

| | A | B | C | D | Total=$T_{i*}$ | $[T_{i*}^2]/k$ | $\Sigma X_{*ij}^2$ |
|---|---|---|---|---|---|---|---|
| 1 | -2 | -8 | 1 | -10 | **-19** | 90.25 | 169 |
| 2 | 0 | -6 | 6 | -3 | **-3** | 2.25 | 81 |
| 3 | -12 | -10 | -2 | -14 | **-38** | 361 | 444 |
| 4 | -3 | -8 | 0 | -13 | **-24** | 144 | 242 |
| 5 | -8 | -4 | 3 | -7 | **-16** | 64 | 138 |
| **Total=$T_{*j}$** | **-25** | **-36** | **8** | **-47** | **-100** | 661.5 | 1074 |
| $[T_{*j}^2]/h$ | 125 | 259.2 | 12.8 | 441.8 | 838.8 | | |

T=Grand Total = -100

Correction Factor = $\dfrac{(Grand\ \ total)^2}{Total\ \ No\ \ of\ \ Observations} = \dfrac{(-100)^2}{20} = 500$

$$TSS = \sum_i \sum_j X_{ij}^2 - C.F = 1074 - 500 = 574$$

$$SSR = \frac{\sum T_{i*}{}^2}{k} - C.F = 661.5 - 500 = 161.5$$

$$SSC = \frac{\sum T_{*j}^{\ 2}}{h} - C.F = 838.8 - 500 = 338.8$$

SSE = TSS – SSC – SSR = 574 – 161.5 – 338.8 = 73.7

**ANOVA Table**

| Source of Variation | Sum of Squares | Degree of freedom | Mean Square | F- Ratio | $F_{Tab}$Ratio |
|---|---|---|---|---|---|
| Between Rows (Workers) | SSR=161.5 | h - 1= 4 | MSR= 40.375 | $F_R$ = 6.574 | $F_{5\%}(4, 12)$ = 3.26 |
| Between Columns (Machine) | SSC=338.8 | k – 1=3 | MSC = 112.933 | $F_C$ = 18.388 | $F_{5\%}(3, 12)$ = 3.59 |
| Residual | SSE = 73.7 | (h – 1)( k – 1) = 12 | MSE = 6.1417 | | |
| Total | 1074 | | | | |

**Conclusion :** Cal $F_C$< Tab $F_C$ and Cal $F_R$< Tab $F_R$ $\Rightarrow$ There is no significant difference between the Machine types and no significant difference between the Workers

4. **A Company appointments four salesmen A, B, C and D and observes their sales in 3 seasons: summer, winter and monsoon. The figures (in lakhs of Rs.) are given in the following table:**

| Season | Salesman | | | |
|---|---|---|---|---|
| | A | B | C | D |
| Summer | 36 | 36 | 21 | 35 |
| Winter | 28 | 29 | 31 | 32 |
| Monsoon | 26 | 28 | 29 | 29 |

**i) Do the salesmen significantly differ in performance?**
**ii) Is there significant difference between the seasons?**
**Solution:**

**Null Hypothesis $H_0$:** There is no significant difference between the sales in the 3 seasons and also between the sales of the 4 salesmen.

Alternate Hypothesis $H_1$**:** There is a significant difference between the sales in the 3 seasons and also between the sales of the 4 salesmen.

**Test statistic:**
To simplify calculations we deduct 30 from each value

| Seasons | | A $X_1$ | B $X_2$ | C $X_3$ | D $X_4$ | Seasons Total | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $Y_1$ | Summer | 6 | 6 | -9 | 5 | **8** | 36 | 36 | 81 | 25 |
| $Y_2$ | Winter | -2 | -1 | 1 | 2 | **0** | 4 | 1 | 1 | 4 |
| $Y_3$ | Monson | -4 | -2 | -1 | -1 | **-8** | 16 | 4 | 1 | 1 |
| Total | | 0 | 3 | -9 | 6 | **0** | 56 | 41 | 83 | 30 |

Step1: N= Total No of Observations = 12

Step 2: T=Grand Total = 0

Step 3: Correction Factor = $\frac{(\text{Grand total})^2}{\text{Total No of Observations}} = \frac{T^2}{N} = \frac{0^2}{12} = 0$

Step 4: TSS= $\sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 56+41+83+30-0 = 210$

Step 5:

$$SSC = \frac{\left(\sum X_1\right)^2}{N_1} + \frac{\left(\sum X_2\right)^2}{N_1} + \frac{\left(\sum X_3\right)^2}{N_1} + \frac{\left(\sum X_4\right)^2}{N_1} - C.F = \frac{0^2}{3} + \frac{3^2}{3} + \frac{(-9)^2}{3} + \frac{6^2}{3} - 0$$

$$SSC = 0+3+27+12-0 = 42$$

Where $N_1$ = Number of elements in each column

Step 6:

$$SSR = \frac{\left(\sum Y_1\right)^2}{N_2} + \frac{\left(\sum Y_2\right)^2}{N_2} + \frac{\left(\sum Y_3\right)^2}{N_2} - C.F = \frac{8^2}{4} + \frac{0^2}{4} + \frac{(-8)^2}{4} + \frac{6^2}{4} - 0 = 16+0+16-0 = 32$$

Where $N_2$ = Number of elements in each row

Step 7: SSE=TSS-SSC-SSR $= 210-42-32$

Step 8: **ANOVA TABLE:**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Sum of Squares | varience | F – ratio |
|---|---|---|---|---|---|
| Between Columns (Salesmen) | SSC=42 | c-1=4-1=3 | $MSC = \frac{SSC}{c-1} = \frac{42}{3} = 14$ | $MSC = \frac{MSE}{MSC} = \frac{22.67}{14} = 1.619$ | $F_C(6,3) = 8.94$ |
| Between rows (Seasons) | SSR =32 | r-1=3-1=2 | $MSR = \frac{SSR}{r-1} = \frac{32}{2} = 16$ | $MSR = \frac{MSE}{MSR} = \frac{22.67}{16} = 1.417$ | $F_R(6,2) = 8.94$ |

| | | | | | |
|---|---|---|---|---|---|
| Error | SSE=136 | N-c-r +1=6 | $MSE = \frac{SSE}{N-c-r+1} = \frac{136}{6} = 22.67$ | | |
| Total | 210 | 11 | | | |

Table Value of F = $F_C(\text{Error}, \text{d.f}) = F_C(6,3) = 8.94$ , $F_R(\text{Error}, \text{d.f}) = 8.94$ with 5% level of significance

**Conclusion:**

**1) Cal $F_R$ < Table $F_{R,0.05}(6,3)$**

Hence we accept the $H_0$ and we conclude that there is no significant difference between sales in the three seasons.

2) **Cal $F_R$ < Table $F_{R,0.05}(6,2)$** .

Hence we accept the $H_0$ and we conclude that there is no significant difference between in the sales of 4 salesmen.

5. **Analyze $2^2$ factorial experiments for the following table.**

| Treatment | Replications | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| (1) | 12 | 12.3 | 11.8 | 11.6 |
| a | 12.8 | 12.6 | 13.7 | 14 |
| b | 11.5 | 11.9 | 12.6 | 11.8 |
| ab | 14.2 | 14.5 | 14.4 | 15 |

**SOLUTION:**

**Null hypothesis: All the mean effects are equal.**

**Let A and B be the two factors.**

**Let n=number of replications=4**

**Subtract 12 from each**

| Treatment | Replications | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| (1) | 0 | 0.3 | -0.2 | -0.4 |
| a | 0.8 | 0.6 | 1.7 | 2 |
| b | -0.5 | -0.1 | 0.6 | -0.2 |
| ab | 2.2 | 2.5 | 2.4 | 3 |

**Let us find SS for the table**

| **Treatment** | **Replications** | | | | **Row Total $R_i$** | $R_i^2$ |
|---|---|---|---|---|---|---|
| | **I** | **II** | **III** | **IV** | | |
| **(1)** | **0** | **0.3** | **-0.2** | **-0.4** | **-0.3** | **0.09** |
| **a** | **0.8** | **0.6** | **1.7** | **2** | **5.1** | **26.01** |
| **b** | **-0.5** | **-0.1** | **0.6** | **-0.2** | **-0.2** | **0.04** |
| **ab** | **2.2** | **2.5** | **2.4** | **3** | **10.1** | **102.01** |
| **Column Total $C_j$** | **2.5** | **3.3** | **4.5** | **4.4** | **T=14.7** | |
| $C_j^2$ | **6.25** | **10.89** | **20.25** | **19.36** | | |

**T=14.7**

**Correction factor=** $\frac{T^2}{N}$ **=13.5**

**TSS=21.19**

**SSC=0.688**

**SSR=18.54**

**SSE=1.962**

| **Source of Variation** | **Sum of Squares** | **Degree of freedom** | **Mean Square** | **F- Ratio** | **$F_{Tab}$ Ratio** |
|---|---|---|---|---|---|
| b | $S_B$=1.63 | 1 | MSB=1.63 | $F_B = 7.409$ | 10.56 |
| a | $S_A$=15.41 | 1 | MSA=15.41 | $F_A = 70.04$ | 10.56 |

| | | | | | |
|---|---|---|---|---|---|
| ab | $S_{AB}$ =1.50 | 1 | MSAB=1.50 | $F_{AB}=6.81$ | 10.56 |
| Error | SSE=1.962 | N-C-r+1=9 | SSE=1.962 | | |

**Cal( $F_A$ )=70.04 $\Rightarrow H_0$ is rejected at 1% level**

**Cal( $F_B$ )=7.409 $\Rightarrow H_0$ is accepted at 1% level**

**Cal( $F_{AB}$ )=10.56 $\Rightarrow H_0$ is accepted at 1% level**

6. **Analyse the variance in the following latin square of yields (in kgs) of paddy where A, B, C, D denote the different methods of cultivation.**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **D** | **122** | **A** | **121** | **C** | **123** | **B** | **122** |
| **B** | **124** | **C** | **123** | **A** | **122** | **D** | **125** |
| **A** | **120** | **B** | **119** | **D** | **120** | **C** | **121** |
| **C** | **122** | **D** | **123** | **B** | **121** | **A** | **122** |

**Examine whether the different methods of cultivation have given significantly different yields.**

**Solution:**

We shift the origin $X_{ij} = x_{ij} - 100$; n = 4; N = 16

| | **I** | **II** | **III** | **IV** | **Total=$T_{i*}$** | **$[T_{i*}^2]/n$** | **$\Sigma X_{*ij}^2$** |
|---|---|---|---|---|---|---|---|
| A | 2 | 1 | 3 | 2 | **8** | 16 | 18 |
| B | 4 | 3 | 2 | 5 | **14** | 49 | 54 |
| C | 0 | -1 | 0 | 1 | **0** | 0 | 2 |
| D | 2 | 3 | 1 | 2 | **8** | 16 | 18 |
| **Total=$T_{*j}$** | **8** | **6** | **6** | **10** | **30** | 81 | 92 |
| $[T_{*j}^2]/n$ | 16 | 9 | 9 | 25 | 59 | | |
| **$\Sigma X_{i*}^2$** | 24 | 20 | 14 | 34 | 92 | | |

| | **Letters** | | | | **Total=$T_{i*}$** | **[$T_{i*}^2$]/n** |
|---|---|---|---|---|---|---|
| P | 1 | 2 | 0 | 2 | **5** | 6.25 |
| Q | 2 | 4 | -1 | 1 | **6** | 9 |
| R | 3 | 3 | 1 | 2 | **9** | 20.25 |
| S | 2 | 5 | 0 | 3 | **10** | 25 |
| **Total** | | | | | **30** | 60.5 |

T=Grand Total = 30 ;Correction Factor = $\frac{(Grand\ \ total)^2}{Total\ \ No\ \ of\ \ Observations} = \frac{(30)^2}{16}$

$$TSS = \sum_i \sum_j X_{ij}^2 - C.F = 92 - \frac{(30)^2}{16} = 35.75$$

$$SSR = \frac{\sum T_{i*}^{\ 2}}{n} - C.F = 81 - \frac{(30)^2}{16} = 24.75$$

$$SSC = \frac{\sum T_{*j}^{\ 2}}{n} - C.F = 59 - \frac{(30)^2}{16} = 2.75$$

$$SSL = \frac{\sum T_{i*}^{\ 2}}{n} - C.F = 60.5 - \frac{(30)^2}{16} = 4.25$$

SSE = TSS – SSC – SSR-SSL = 35.75 – 24.75 – 2.75 – 4.25 = 4

**ANOVA Table**

| **Source of Variation** | **Sum of Squares** | **Degree of freedom** | **Mean Square** | **F- Ratio** | **$F_{Tab}$Ratio ( 5% level)** |
|---|---|---|---|---|---|
| Between Rows | SSR=24.75 | n - 1= 3 | MSR=8.25 | $F_R$= 12.31 | $F_R$(3, 6)=4.76 |
| Between Columns | SSC=2.75 | n - 1= 3 | MSC = 0.92 | $F_C$ = 1.37 | Fc(3, 6)=4 .76 |
| Between Letters | SSL = 4.25 | n - 1= 3 | MSL = 1.42 | $F_L$ = 2.12 | |
| Residual | SSE= 4 | (n – 1)(n – 2) = 6 | MSE = 0.67 | | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | $F_L(3, 6)=4$ .76 |
| Total | 35.75 | | | | |

**Conclusion :**

Cal $F_C$< Tab $F_C$ , Cal $F_L$< Tab $F_L$ and Cal $F_R$> Tab $F_R$ $\Rightarrow$ There is significant difference between the **rows** , no significant difference between the letters and no significant difference between the columns

7. **A variable trial was conducted on wheat with 4 varieties in a Latin Square Design. The plan of the experiment and the per plot yield are given below :**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **C** | **25** | **B** | **23** | **A** | **20** | **D** | **20** |
| **A** | **19** | **D** | **19** | **C** | **21** | **B** | **18** |
| **B** | **19** | **A** | **14** | **D** | **17** | **C** | **20** |
| **D** | **17** | **C** | **20** | **B** | **21** | **A** | **15** |

**Analyse data and interpret the result.**

$H_0$ : Four varieties are similar

$H_1$ : Four varieties are not similar

Let us take 20 as origin for simplifying the calculation

| Variety | $X_1$ | $X_2$ | $X_3$ | $X_4$ | TOTAL | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|---|
| $Y_1$ | 5 | 3 | 0 | 0 | **8** | 25 | 9 | 0 | 0 |
| $Y_2$ | -1 | -1 | 1 | -2 | **-3** | 1 | 1 | 1 | 4 |
| $Y_3$ | -1 | -6 | -3 | 0 | **-10** | 1 | 36 | 9 | 0 |
| $Y_4$ | -3 | 0 | 1 | -5 | **-7** | 1 | 0 | 1 | 25 |
| | **0** | **-4** | **-1** | **-7** | **-12** | **9** | **46** | **11** | **29** |

N=Total No of Observations = 16 T=Grand Total = -12

$$\text{Correction Factor} = \frac{(\text{Grand total})^2}{\text{Total No of Observations}} = 9$$

$$TSS = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 - C.F = 36 + 46 + 11 + 29 - 9 = 113$$

$$SSC = \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - C.F = \frac{(6)^2}{4} + \frac{(10)^2}{4} + \frac{(6)^2}{4} + \frac{(10)^2}{4} - 9 = 4$$

$$SSR = \frac{(\sum Y_1)^2}{N_1} + \frac{(\sum Y_2)^2}{N_2} + \frac{(\sum Y_3)^2}{N_2} + \frac{(\sum Y_4)^2}{N_2} - C.F = \frac{(8)^2}{4} + \frac{(-3)^2}{4} + \frac{(-10)^2}{4} + \frac{(-7)^2}{4} - 9 = 46.5$$

To find SSK

| Treatment | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| A | 0 | -1 | -6 | -5 | -12 |
| B | 3 | -2 | -1 | 1 | 1 |
| C | 5 | 1 | 0 | 0 | 6 |
| D | 0 | -1 | -3 | -3 | -7 |

SSK= $\frac{(\sum Y_1)^2}{K_1} + \frac{(\sum Y_2)^2}{K_2} + \frac{(\sum Y_3)^2}{K_3} + \frac{(\sum Y_4)^2}{K_4} - C.F$

$= 48.5$

SSE= TSS − SSC−SSR−SSK = 113-7.5-46.5-48.5=10.5

**ANOVA Table**

| **Source of Variation** | **Sum of Squares** | **Degree of freedom** | **Mean Square** | **F- Ratio** |
|---|---|---|---|---|
| Column Treatment | SSC=7.5 | n-1=3 | $MSC = \frac{SSC}{n-1}$ =2.5 | $F_C = \frac{MSC}{MSE}$=1.43 |
| Row Treatments | SSR=46.5 | n-1=3 | $MSR = \frac{SSR}{n-1}$ =15.5 | $F_R = \frac{MSE}{MSR}$=8.86 |
| Between Treatments | SST=48.5 | n-1=3 | $MSK = \frac{SSK}{n-1}$ =16.17 | $F_K = \frac{MSK}{MSE}$=9.24 |
| Error (or) Residual | SSE=10.5 | (n-1) (n-2)=6 | $MSE = \frac{SSE}{(n-1)(n-2)} = 1.75$ | |

Table value F(3,6) degrees of freedom 8.94

There is significant difference between treatments

## STATISTICAL QUALITY CONTROL

**Statistical quality control:**

**statistical quality control is a statistical method for finding whether the variation in the quality of the product is due to random causes or assignable causes.**

**Objectives of statistical quality control:**

**To achieve better utilization of raw materials, to control waste and scrap and to optimize the quality of the product without any defects.**

**Control chart:**

**It is a useful graphical method to find whether a process is in statistical quality control.**

**Uses of Quality control chart:**

**It helps in determining whether the goal set is being achieved by finding out whether the Process is in control or not.**

**Different types of control chart:**

**Control chart for variables – Range and mean chart, Control chart for attributes- p-chart,**

**C-chart, np-chart.**

**control limits for mean chart:**

**Central limit =** $\bar{\bar{x}}$ **, upper control limit =** $\bar{\bar{x}} + A_2\bar{R}$ **, lower control limit =** $\bar{\bar{x}} - A_2\bar{R}$

**Where** $\bar{x}$ **is the mean of the sample and R is the range.**

**The control limits for range chart:**

**CL=** $\bar{R}$ **, UCL =** $D_4\bar{R}$ **, LCL=**$D_3\bar{R}$ **.**

**Procedure to draw the** $\bar{x}$**−chart & R-chart:**

**1. The sample values in each of the N samples each of size 'n' will be given. Let** $\overline{X_1}, \overline{X_2}, \ldots \overline{X_N}$ **be the means of the N samples &** $R_1, R_2 \ldots R_N$ **be the ranges of the N samples.**

**2. Compute** $\overline{\overline{X}} = \frac{1}{N}\left(\overline{X_1} + \overline{X_2} + \ldots + \overline{X_N}\right); \overline{R} = \frac{1}{N}\left(R_1 + R_2 + \ldots + R_N\right)$

**3. The values of** $A_2, D_3, D_4$ **for the given sample size n are taken from the table of control chart constants.**

**4. Find the values of the control limits** $\bar{\bar{x}} \pm A_2\bar{R}$ **( for the mean chart) and the control limits** $D_3\bar{R}$ **and** $D_4\bar{R}$ **( for the range chart) are computed.**

**5. On the ordinary graph sheet, the sample numbers are represented on the x-axis and the sample means on the**

**y-axis ( for the mean chart) and the sample ranges on the y-axis(for the range chart).**

**6.For drawing the mean chart, we draw the three lines** $y = \overline{\overline{X}}$ **,** $y = \overline{\overline{X}} - A_2\overline{R}$ **and** $y = \overline{\overline{X}} + A_2\overline{R}$ **which represent respectively the central line, the L.C.L line and U.C.L line, Also we plot the points whose coordinates are** $\left(1, \overline{X_1}\right), \left(2, \overline{X_2}\right), \ldots \left(N, \overline{X_N}\right)$ **and join adjacent points by line segments. The graph thus obtained is the** $\overline{X}$ **chart.**

**7. For drawing the mean chart, we draw the three lines** $y = \overline{R}$ **,** $y = D_3\overline{R}$ **and** $y = D_4\overline{R}$ **which represent respectively the central line, the L.C.L line and U.C.L line, Also we plot the points whose coordinates are** $(1, R_1), (2, R_2), \ldots (N, R_N)$ **and join adjacent points by line segments. The graph thus obtained is the** $R$ **chart.**

**Mean And Range Chart Problems**

1. **Given below are the values of sample mean** $\overline{X}$ **and sample range R for 10 samples, each of size 5. Draw the appropriate mean and range charts and comment on the state of control on the state of control of the process.**

| Sample No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean $\overline{X}_i$ | 43 | 49 | 37 | 44 | 45 | 37 | 51 | 46 | 43 | 47 |
| Range $R_i$ | 5 | 6 | 5 | 7 | 7 | 4 | 8 | 6 | 4 | 6 |

**Solution:**

$$\overline{\overline{X}} = \frac{1}{N}\sum \overline{X_i}$$
$$= \frac{1}{10}\left[43 + 49 + 37 + 44 + 45 + 37 + 51 + 46 + 43 + 47\right]$$
$$= 44.2$$

$$\overline{R} = \frac{1}{N}\sum R_i$$
$$= \frac{1}{10}\left[5 + 6 + 5 + 7 + 7 + 4 + 8 + 6 + 4 + 6\right]$$
$$= 5.8$$

**From the table of control chart for sample size n=5, we have** $A_2 = 0.577, D_3 = 0 \,\&\, D_4 = 2.115$

i) **Control limits for** $\overline{X}$ **chart:**

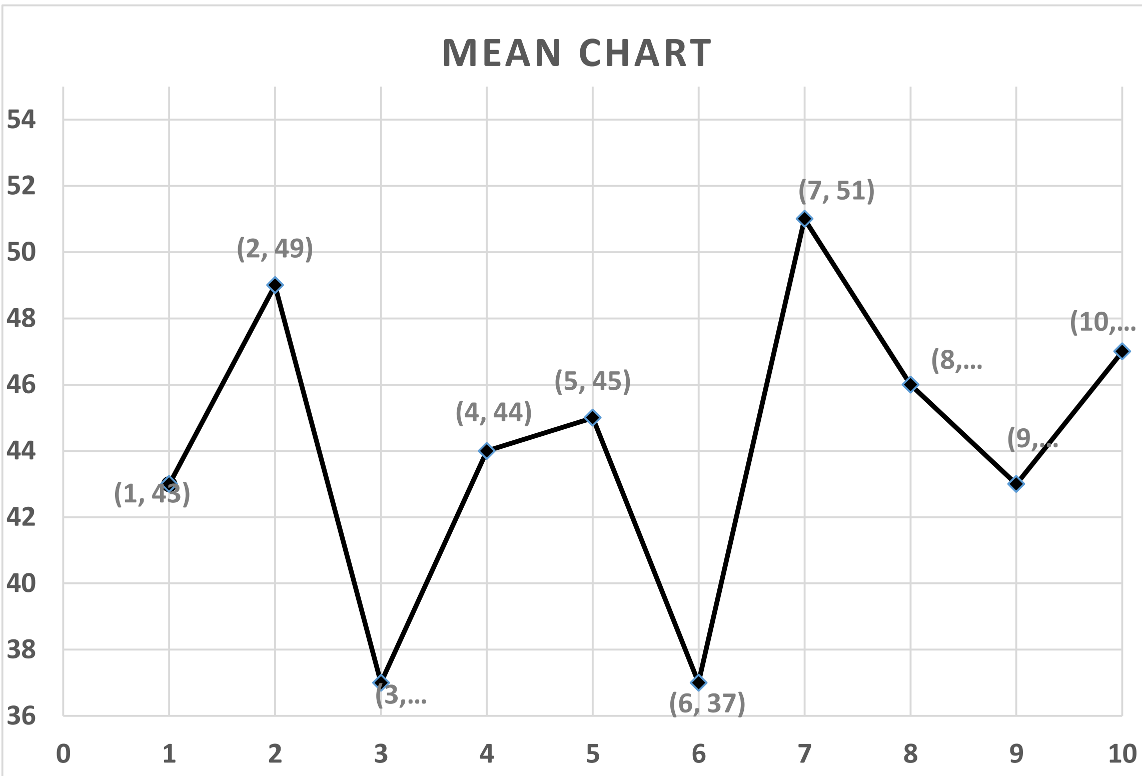
**CL (central line) =** $\overline{\overline{X}} = 44.2$

$$LCL = \overline{\overline{X}} - A_2\overline{R} = 44.2 - (0.577)(5.8) = 40.85$$

$$UCL = \overline{\overline{X}} + A_2\overline{R} = 44.2 + (0.577)(5.8) = 47.55$$

Conclusion :

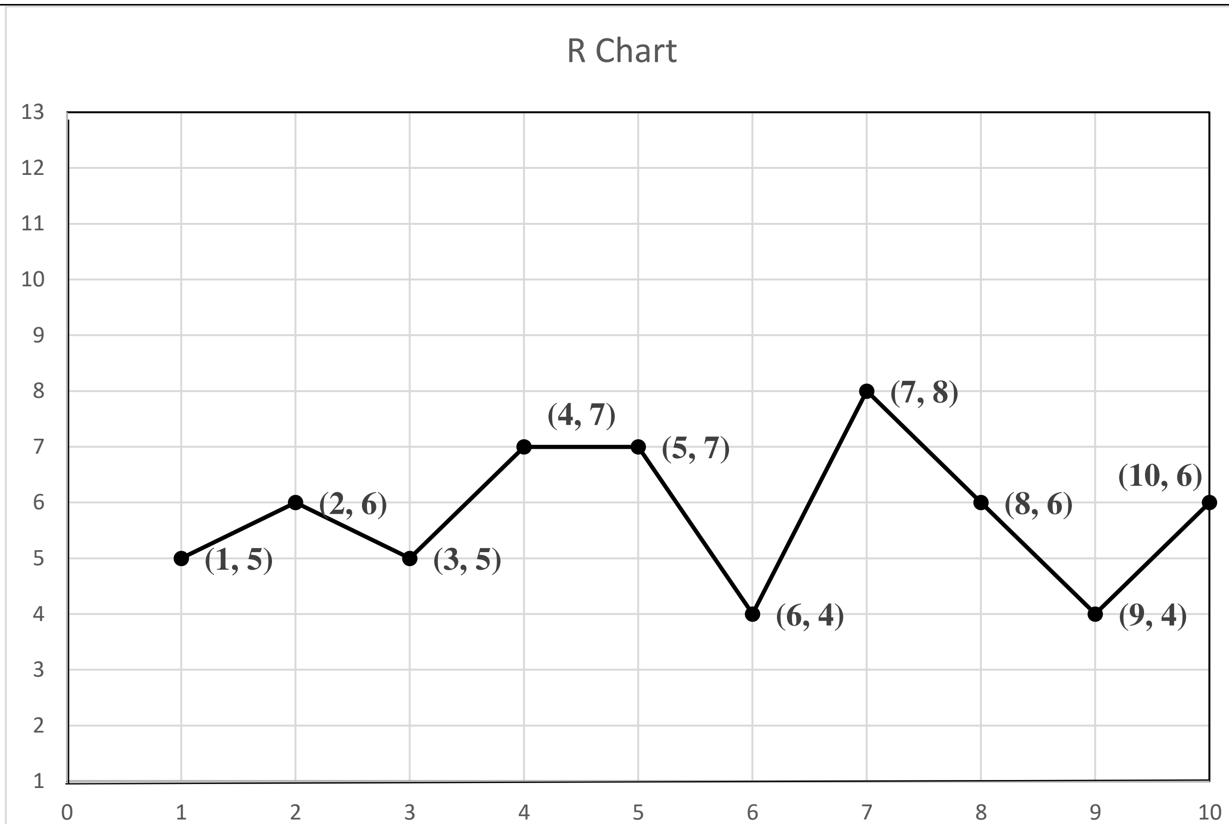
Since $2^{nd}$,$3^{rd}$,$6^{th}$and $7^{th}$ sample means fall outside the control limits the statistical process is out of control according to $\overline{X}$ *chart*




**ii) Control limits for R-Chart:**

$$CL = \overline{R} = 5.8; \;\; LCL = D_3; \;\; \overline{R} = 0$$

$$LCL = D_4\overline{R} = (2.115)(5.8) = 12.267 \approx 12.27$$




**Conclusion :**

Since all the sample means fall within the control limits the statistical process is under control according to $R\,chart$.

2. **The following data give the measurements of 10 samples each of size 5 in the production process taken in an interval of 2 hours. Calculate the sample means and ranges and draw the control charts for mean and range.**

| **Sample No.** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Observed measurements X** | **49** | **50** | **50** | **48** | **47** | **52** | **49** | **55** | **53** | **54** |
| | **55** | **51** | **53** | **53** | **49** | **55** | **49** | **55** | **50** | **54** |
| | **54** | **53** | **48** | **51** | **50** | **47** | **49** | **50** | **54** | **52** |
| | **49** | **46** | **52** | **50** | **44** | **56** | **53** | **53** | **47** | **54** |
| | **53** | **50** | **47** | **53** | **45** | **50** | **45** | **57** | **51** | **56** |

**Solution:**

$$\overline{\overline{X}} = \frac{1}{N}\sum \overline{X_i}$$

$$= \frac{1}{10}\left[52+50+50+51+47+52+49+54+51+54\right]$$

$$= 51.0$$

4

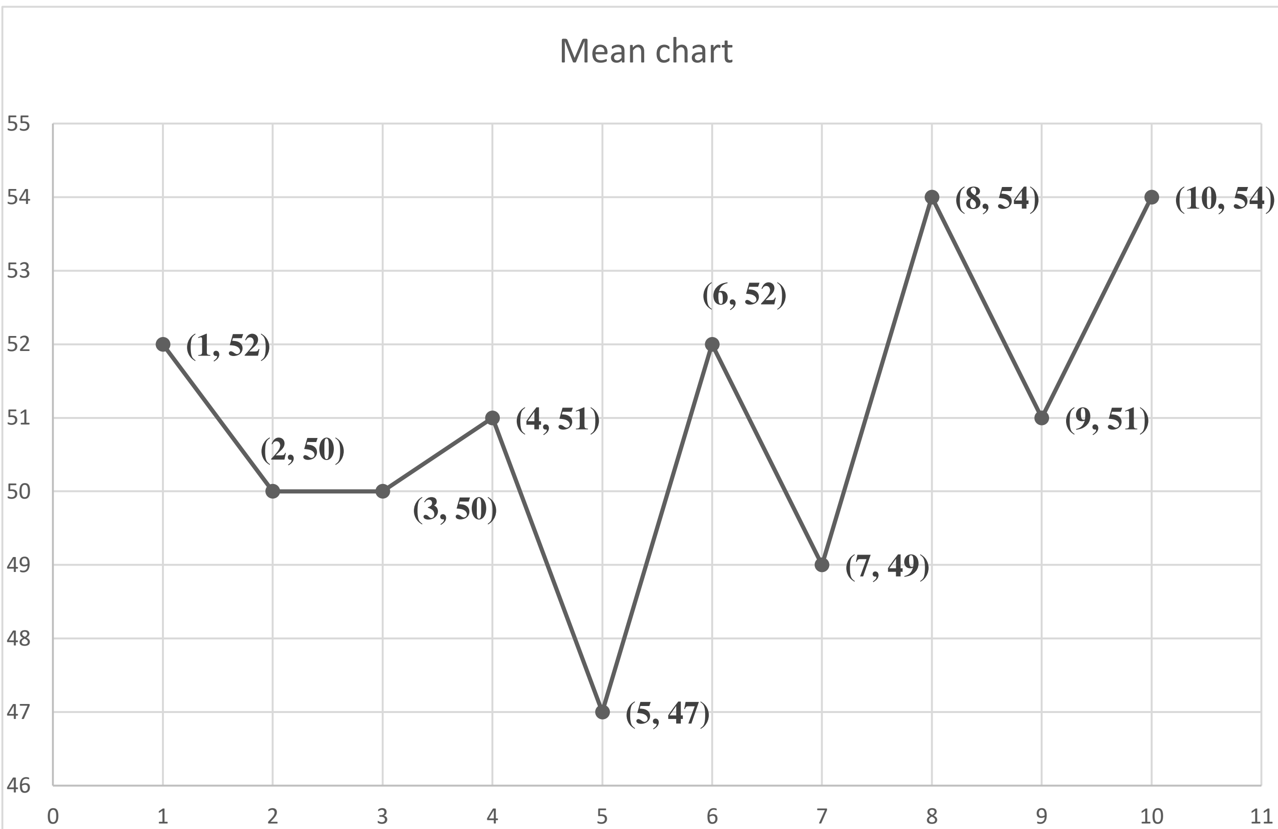
$$\overline{R} = \frac{1}{N}\sum R_i$$
$$= \frac{1}{10}\left[6+7+6+5+6+9+8+7+7+4\right]$$
$$= 6.5$$

**From the table of control chart for sample size n=5, we have** $A_2 = 0.577, D_3 = 0 \;\&\; D_4 = 2.115$

**i) Control limits for $\overline{X}$ chart:**

**CL (central line) =** $\overline{\overline{X}} = 44.2$

$$LCL = \overline{\overline{X}} - A_2\overline{R}_2 = 51.0 - (0.577)(6.5) = 47.2495$$
$$UCL = \overline{\overline{X}} + A_2\overline{R}_2 = 51.0 + (0.577)(6.5) = 54.7505$$
$$CL = \overline{X} = 51.0$$

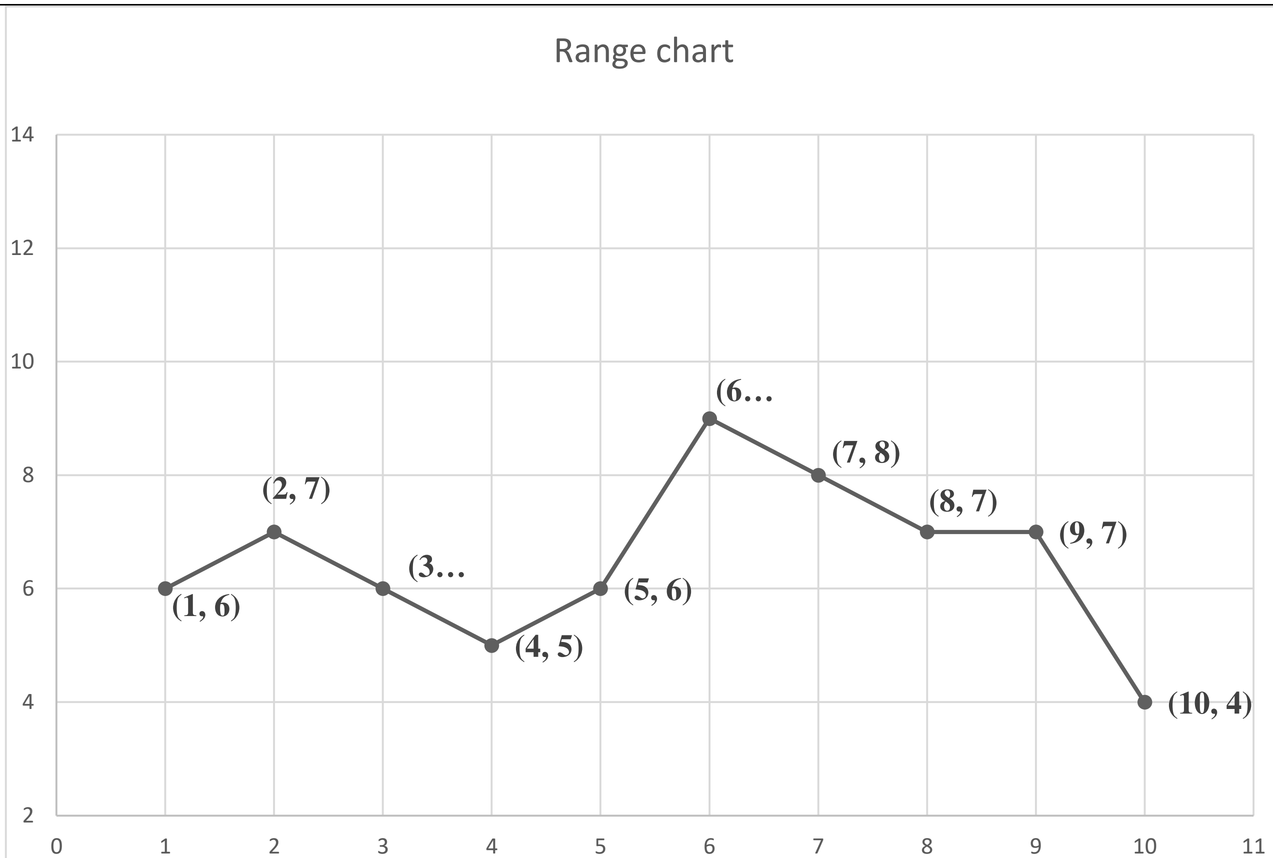



**Conclusion :**

Since 5th sample mean fall outside the control limits the statistical process is out of control according to $\overline{X}$ *chart*

**ii) Control limits for R-Chart:**

$$CL = \overline{R} = 6.5; \quad LCL = D_3\ \overline{R} = 0$$
$$UCL = D_4\overline{R} = (2.115)(6.5) = 13.7475$$




**Conclusion :**

Since all the sample means fall within the control limits the statistical process is under control according to $R\,chart$.

**C-chart:**

**Control chart for number of defects is called c-chart.**

**The control limits for c-chart.**

**CL =** $\bar{c}$ **UCL =** $\bar{c} + 3\sqrt{c}$ **LCL =** $\bar{c} - 3\sqrt{c}$

**C-Chart problems**

1. **15 tape recorders were examined for quality control test. The number of defects in each tape recorder is recorded below. Draw the appropriate control chart and comment on the state of control.**

| Unit No.(i) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of defects (c) | 2 | 4 | 3 | 1 | 1 | 2 | 5 | 3 | 6 | 7 | 3 | 1 | 4 | 2 | 1 |

**Solution:**

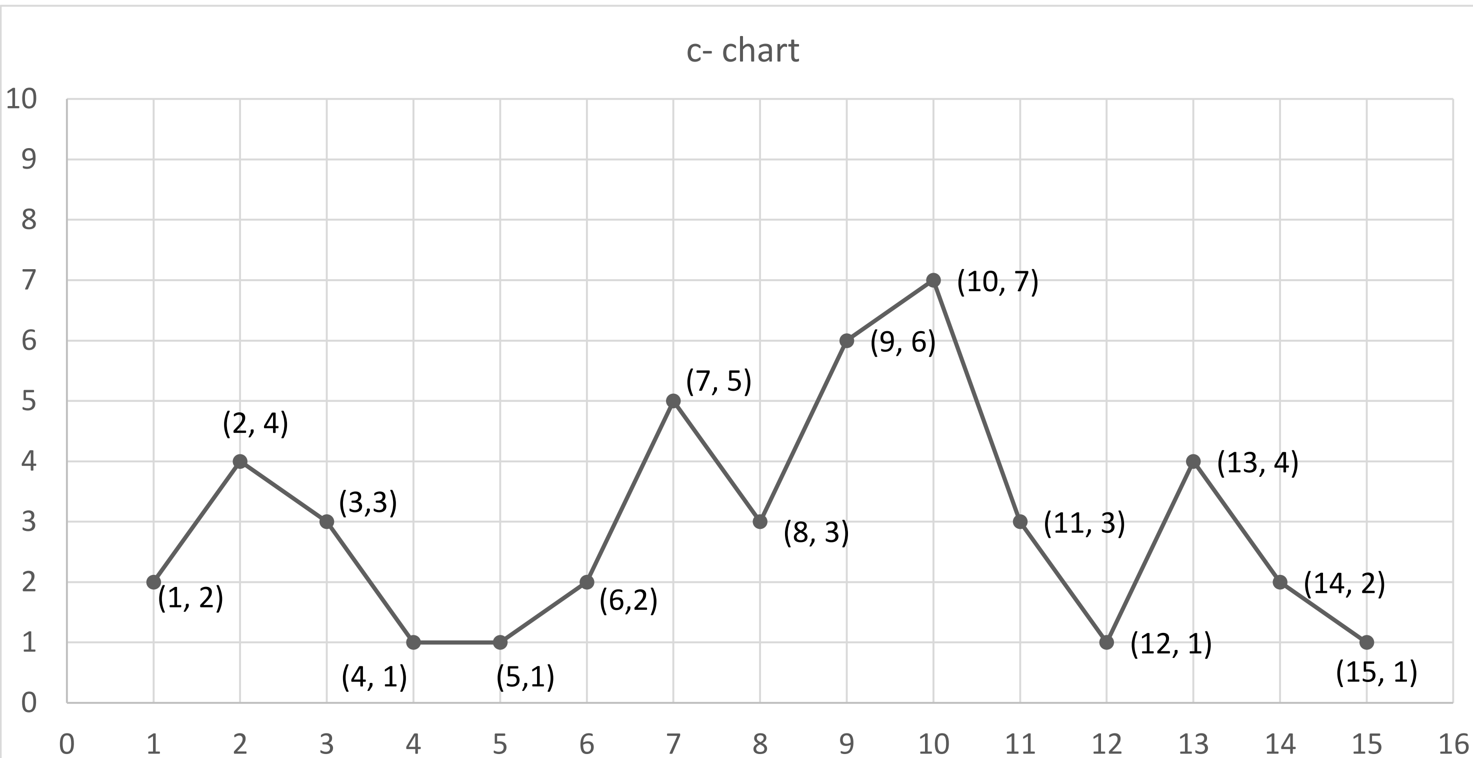
The number of defects per sample containing only one item is given,

$$\bar{c} = \frac{\sum c_i}{N} = \frac{(2+4+3+\cdots+2+1)}{15} = \frac{45}{15} = 3$$

$\mathrm{CL} = \bar{\mathrm{c}} = 3; \quad \mathrm{LCL} = \bar{\mathrm{c}} - 3\sqrt{\mathrm{c}} = 3 - 3\sqrt{3} = -2.20$

We take $\mathrm{LCL} = 0$ ( since LCL cannot be –ve)

$$\mathrm{UCL} = \bar{\mathrm{c}} + 3\sqrt{\mathrm{c}} = 3 + 3\sqrt{3} = 8.20$$




Since all the sample points lie within the LCL and UCL lines, the process is under control.

**2. 20 pieces of cloth out of different rolls contained respectively 1,4,3,2,4,5,6,7,2,3,2,5,7,6,4,5,2,1,3 and 8**

**imperfections.**

**Ascertain whether the process is in a state of statistical control.**

**Solution:**

Let C denote the number of imperfections per unit.

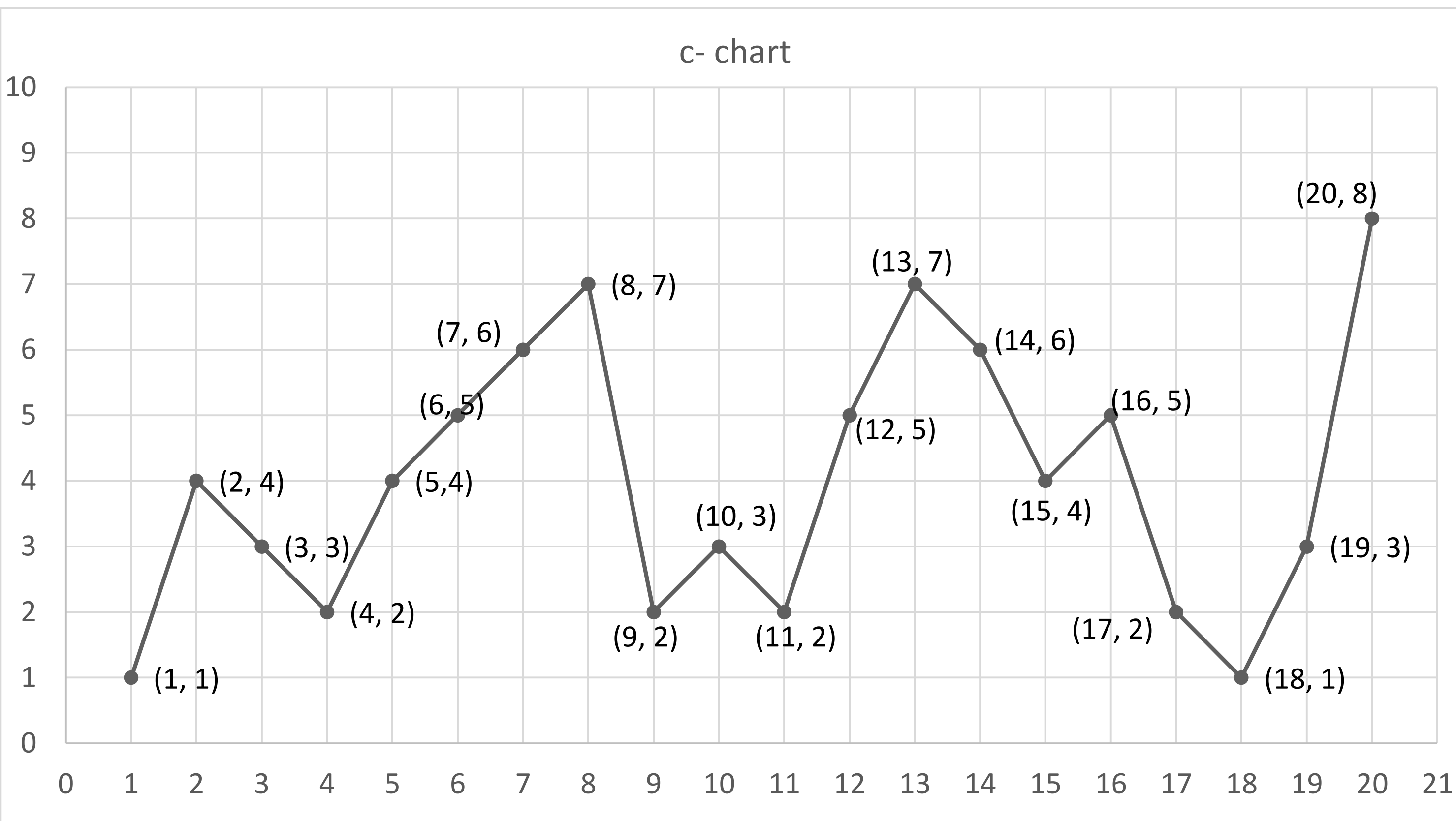
$$\bar{c} = \frac{Total\ no\ of\ defects}{Total\ sample\ inspected} = \frac{\sum c}{n}$$

$$\bar{c} = \frac{1+4+3+2+4+5+\ldots+1+3+8}{20}$$ **=4**

**UCL=** $\bar{C} + 3\sqrt{\bar{C}} = 10$

**LCL=** $\bar{C} - 3\sqrt{\bar{C}} = -2$

We take $LCL = 0$ ( since LCL cannot be –ve)




Since all the sample points lie within the LCL and UCL lines, the process is under control.

**p-chart:**

**Control chart for fraction defectives is called p-chart.**

**control limits for p-chart.**

**UCL=** $n\bar{p} + 3\sqrt{n\bar{p}\bar{q}}$ **, LCL=** $n\bar{p} - 3\sqrt{n\bar{p}\bar{q}}$ **CL =** $n\bar{p}$

**np –chart.**

**Control chart for number of defectives is called np chart.**

**P-Chart & nP-Chart Problems**

1. **Construct a control chart for defectives for the following data:**

| **Sample No:** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **No. inspected :** | **90** | **65** | **85** | **70** | **80** | **80** | **70** | **95** | **90** | **75** |
| **No. of defectives:** | **9** | **7** | **3** | **2** | **9** | **5** | **3** | **9** | **6** | **7** |

**Solution:**

We note that the size of the simple varies from sample to sample. We can construct P-chart, provided $0.75\,\bar{n} < n_i < 1.25\bar{n}$, for all i.

Here

$$\bar{n} = \frac{1}{N}\sum n_i = \frac{1}{10}(90+65+.....+90+75)$$

$$= \frac{1}{10}(800) = 80$$

Hence The values of $n_i$ be between 60 and 100. Hence p-chart can be drawn by the method given below. Now $\bar{p} = \frac{Total\ no.of\ defectives}{Total\ no.of\ items\ inspected}$

$$= \frac{60}{800} = 0.075$$

Hence for the p-chart to be constructed, CL= $\bar{p}$ =0.075

LCL= $\bar{p}-3\sqrt{\dfrac{\bar{p}\left(1-\bar{p}\right)}{\bar{n}}}$

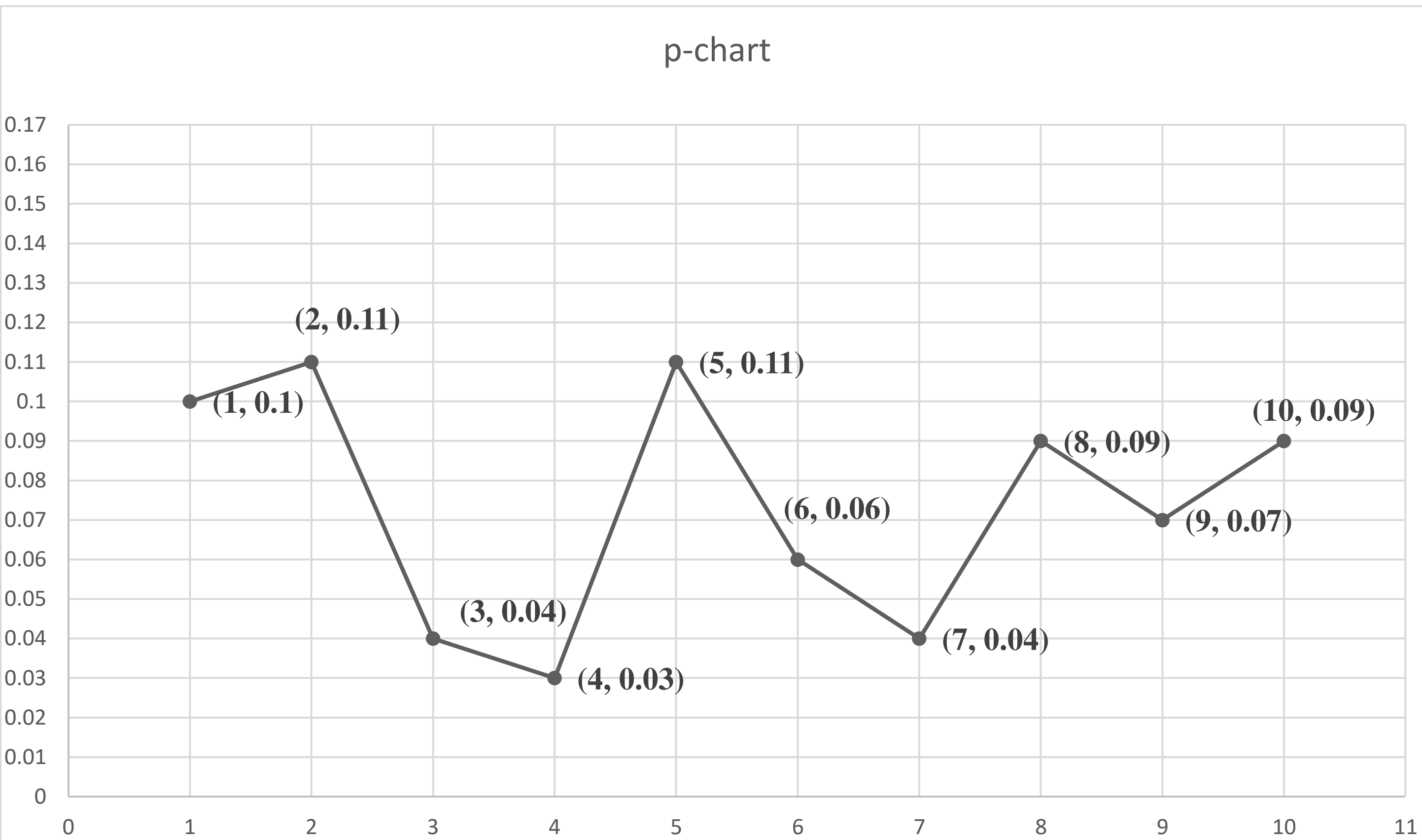
$$=0.075-3\sqrt{\frac{0.075\,X\,0.925}{80}}=-0.013$$

Since LCL cannot be negative,it is taken 0.

UCL= $\bar{p}+3\sqrt{\dfrac{\bar{p}\left(1-\bar{p}\right)}{\bar{n}}}$

$$=0.075+3\sqrt{\frac{0.075\,X\,0.925}{80}}=0.163$$

The values of $p_i$ for the various samples are 0.100, 0.108, 0.035, 0.029, 0.113, 0.063, 0.043, 0.095, 0.067, 0.093




Since all the sample points lie within the control lines, the process is under control.

2. **The following are the figures for the number of defectives of 10 samples each containing 100 items 8,10,9,8,10,11,7,9,6,12 .Draw control chart for fraction defective and comment on the state of control of the process.**

**P for sample = ** $= \frac{No.of\ defectives\, in\, the\, sample}{No.of\ items\, in\, the\, sample}$

**P for sample =** $\frac{8}{100} = 0.08$

**Similarly calculate p for each sample and tabulate.Divide the number of defectives by 100 to get the fraction defective.**

| **Sample No:** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **No. of defectives:** | **8** | **10** | **9** | **8** | **10** | **11** | **7** | **9** | **6** | **12** |
| **P=fraction defectives** | **0.08** | **0.10** | **0.09** | **0.08** | **0.10** | **0.11** | **0.07** | **0.09** | **0.06** | **0.12** |

$$\overline{p} = \frac{\sum p}{n}$$

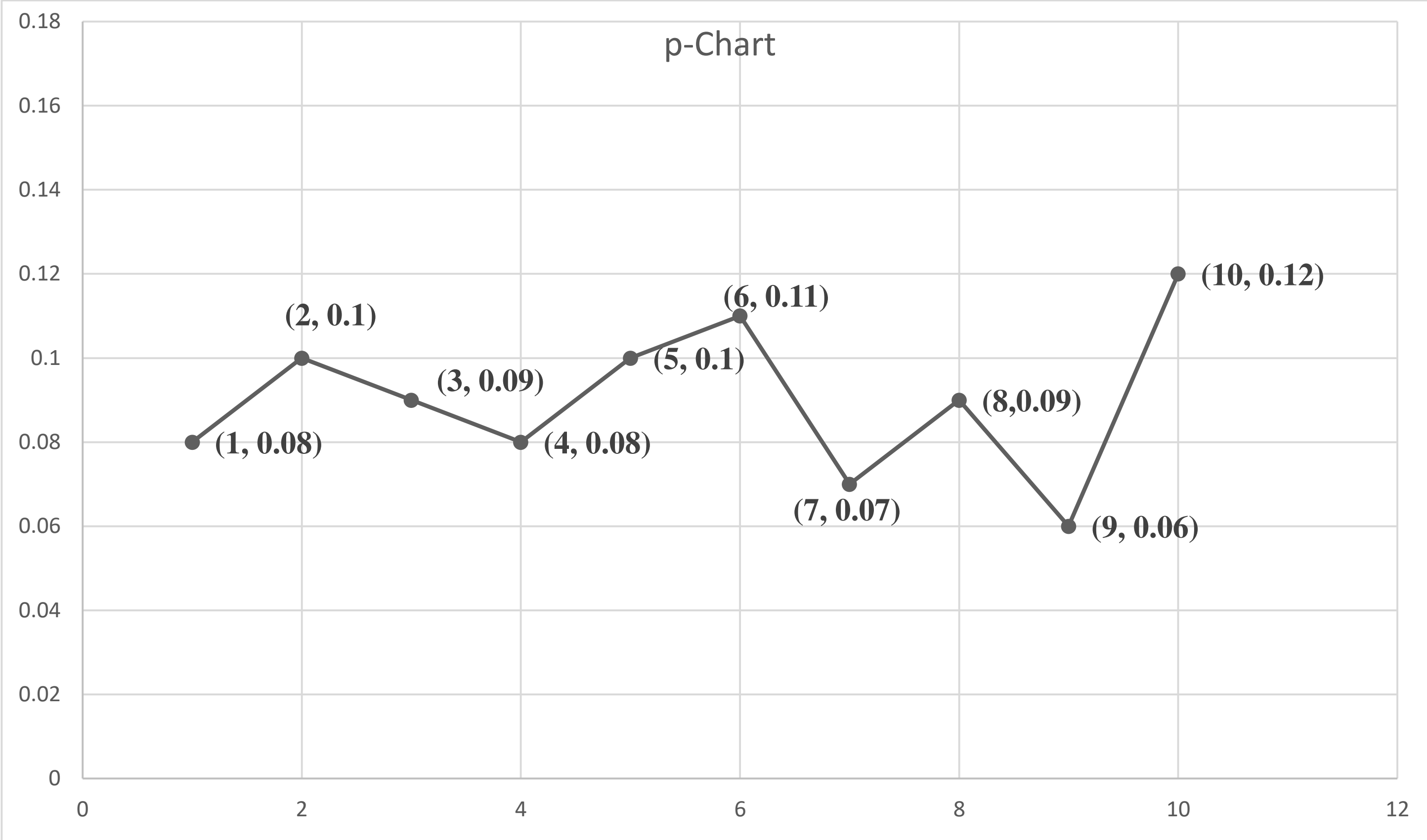
$$= \frac{0.08 + 0.10 + 0.09 + 0.08 + 0.10 + 0.11 + 0.07 + 0.09 + 0.06 + 0.12}{10} = 0.09$$

**UCL=** $\overline{P} + 3\sqrt{\frac{\overline{P}\left(1 - \overline{P}\right)}{n}}$

$= 0.09 + 3\sqrt{\frac{0.09(0.91)}{100}} = 0.177$

**UCL=** $\overline{P} - 3\sqrt{\frac{\overline{P}\left(1 - \overline{P}\right)}{n}}$

$= 0.09 - 3\sqrt{\frac{0.09(0.91)}{100}} = 0.003$




Since all the sample points lie within the control lines, the process is under control.

3. **The data given below are the number of defectives in 10 samples of 100 items each. Construct a p-chart and an np-chart and comment on the results.**

| **Sample No.** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **No. of defectives** | **6** | **16** | **7** | **3** | **8** | **12** | **7** | **11** | **11** | **4** |

**Solution:**

Sample size is constant for all samples, n=100.

Total no. of defectives = 6 + 16+7+3+8+12+7+11+11+4= 85

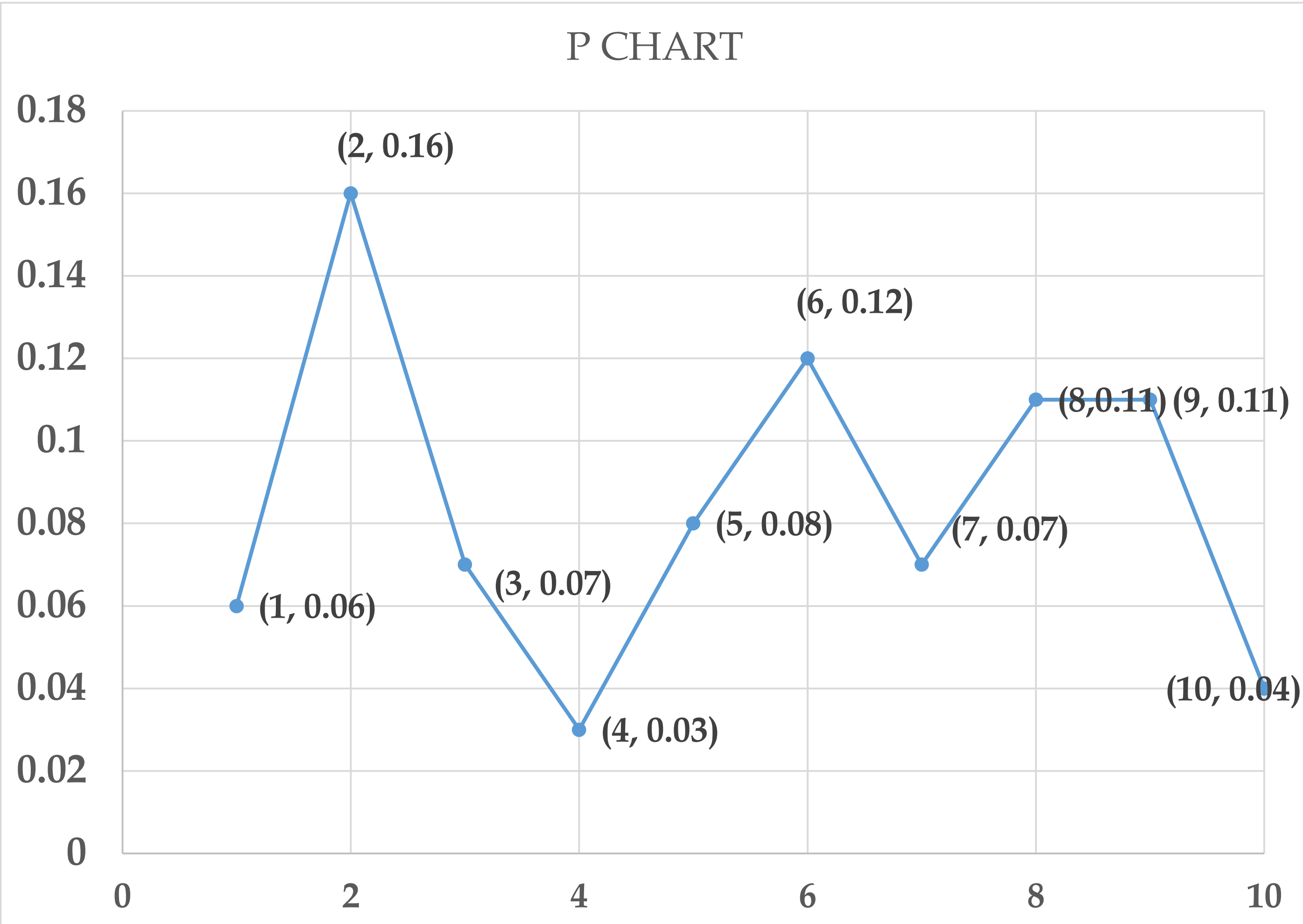
Total no. Inspected= 10 x 100 = 1000

Average fraction defective = $\overline{\mathrm{p}} = \frac{\text{Total no. of defectives}}{\text{Total no. of items inspected}} = \frac{85}{1000} = 0.085$

**For p-chart:**

$$LCL = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.085 - 3\sqrt{\frac{(0.085)(0.915)}{100}} = 0.0013$$

$$UCL = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.085 + \left(\sqrt{\frac{(0.085)(0.915)}{3}}\right) = 0.1687$$




**Conclusion:**

**All these values are less than UCL=0.1687 and greater than LCL=0.0013. In the control chart, all sample points lie within the control limits. Hence, the process is under statistical control.**

**For np-chart:**

$$UCL = n\overline{p} + 3\sqrt{n\overline{p}\left(1-\overline{p}\right)}$$

$$= n\left[\overline{p} + 3\sqrt{\frac{\overline{p}\left(1-\overline{p}\right)}{n}}\right]$$

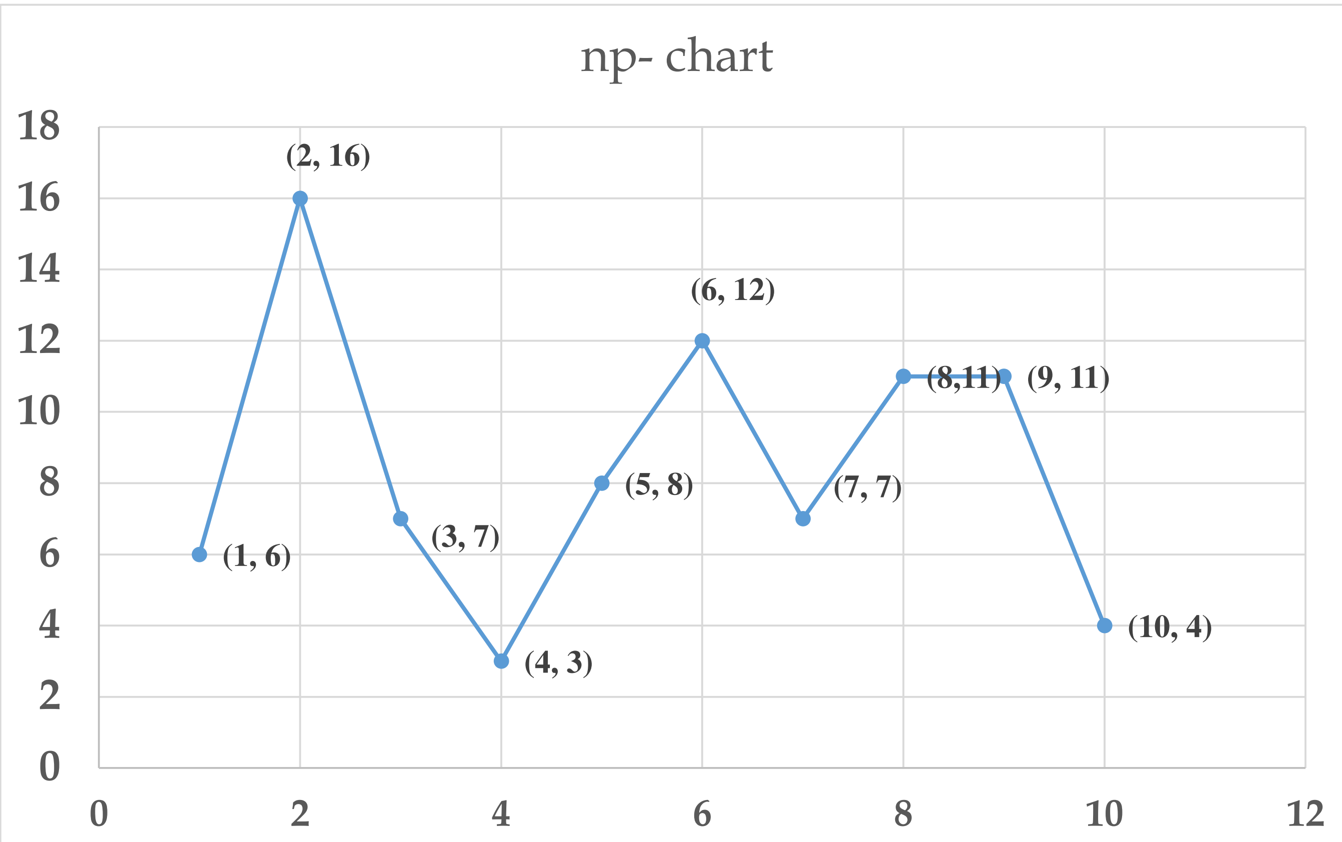
$$= 100(0.1687) = 16.87$$

$$n\bar{p} = 100(0.085) = 8.5$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$$

$$= n\left[\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

$$= 100(0.0013) = 0.13$$




**Conclusion:**

**All the values of number of defectives in the table lie between 16.87 and 0.13. Hence, the process is under control even in np-chart.**