Pimpri Chinchwad Education Trust's

Pimpri Chinchwad College of Engineering

Department of Civil Engineering

ENGINEERING MATHEMATICS III

UNIT III

CHAPTER 2 : PROBABILITY

Mr. Dinesh Kute

(7709495376)

Introduction

3.6.1 Introduction :

Probability theory is the mathematical modeling of the phenomenon of chance or randomness. If a coin is tossed in a random manner, it can land heads or tails, but we do not know which of these will occur on a single toss. However, suppose we let S be the number of times heads appears when the coin is tossed n times. As n increases, the ratio $f = \frac{S}{n}$, called the relative frequency of the outcome, becomes more stable. If the coin is perfectly balanced, then we expect that the coin will land heads approximately 50% of the time or, in other words, the relative frequency will approach $\frac{1}{2}$. Alternatively, assuming the coin is as likely to occur as the other; hence chances of getting a head is one in two, which means the probability of getting is $\frac{1}{2}$. Although the specific outcome on any one toss is unknown, the behavior over the long run is determined. This stable long–run behavior of random phenomena forms the basis of probability theory.

Types of Probability

Theory of Probability

A) Terminology :

i) Experiment :

The performance or action with a view of some interest is called an experiment.

ii) Random experiment :

An experiment which can result in two or more outcomes and the outcome at a particular performance of the experiment is uncertain is called random experiment.

iii) Sample space :

The set of all possible outcomes of an experiment is called a sample space. Every element of sample space is a **sample point**. The sample space denoted by letter 'S' and n(S) represents the total number of sample points total number of sample points in a sample space.

iv) Event :

Any subset of sample space associated with a random experiment is called an event.

Types of Probability

v) Complementary Event :

Consider a two events $A = \{1, 2, 3\}$ $B = \{4, 5, 6\}$ from the sample space $S = \{1, 2, 3, 4, 5, 6\}$

then

A^C = complement of set $A = \{4, 5, 6) = B$

B^C = complement of set $B = \{1, 2, 3\} = A$

Here set A and B are called complementary events.

vi) Mutually Exclusive Events :

Two events A and B are mutually exclusive if $A \cap B = \phi$, that is empty set; there is no common point in A & B.

vii) Exhaustive Events :

Two events A and B are mutually exhaustive if $A \cup B = S$.

i.e. $A \cup B$ contains all sample points.

Definition of Probability

viii) Independent Events :

Two events are independent if occurrence of one event is not affected by the occurrence of the other.

B) Definition of Probability :

If A is an event from an experiment having 'n' mutually exclusive exhaustive and equally likely outcomes out of which 'm' are favourable to an event A, then the probability of event A denoted by P(A) is given by,

$$P(A) = \frac{\text{Number of sample points in A}}{\text{Number of sample points in S}} = \frac{n(A)}{n(S)} = \frac{m}{n}$$

If $n(A) = 0$, then $P(A) = \frac{0}{n} = 0$ and

If $n(A) = n(S)$, then $P(A) = \frac{n(S)}{n(S)} = 1$

Thus we have $0 \leq P(A) \leq 1$ i.e. the probability lies between 0 and 1.

Examples on Probability

Example : 1

From a class of 12 students, 5 are boys and rests are girls. Find the probability that a student selected is a girl.

Solution :

$$n(S) = 12$$

Let A = Event 'a student selected is a girl'.

$$\therefore \; n(A) = {}^7C_1 = 7$$

$$\therefore \; P(A) = \frac{n(A)}{n(S)} = \frac{7}{12}$$

Example : 2

Find the probability of drawing the king from a well shuffled pack of cards.

Solution : $n(S)$ = Total number of cards = 52

Let, A = Event 'a card draw is a king'.

One king card can be drawn from 4 king cards in ${}^4C_1 = 4$ ways.

$$\therefore \; n(A) = {}^4C_1 = 4$$

$$\therefore \; \text{Required probability} = \frac{n(A)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

Examples on Probability

Example : 3

A throw is made with two dice. Find the probability of getting a score of :

i) 10 points

ii) At least 10 points

iii) At most 10 points.

Solution : Sample space : set of all possible outcomes

$$S = \left\{ \begin{array}{l} (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6) \\ (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6) \\ (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6) \\ (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6) \\ (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6) \\ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \end{array} \right\}$$

$$\therefore \quad n(S) = 36$$

i) Let A_1 = Event 'getting a score of 10 points'

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

$$\Rightarrow \quad n(A_1) = 3$$

$$P\text{ (Score of 10 points)} = \frac{n(A_1)}{n(S)} = \frac{3}{36} = \frac{1}{12}$$

Examples on Probability

ii) Let A_2 = Event 'getting at least 10 points'

i.e. getting score of 10, 11 & 12

$$A_2 = \{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}$$

$$\Rightarrow \quad n(A_2) = 6$$

$$P\text{ (Score of at least 10 points)} = \frac{6}{36} = \frac{1}{6}$$

iii) A_3 = Event 'getting score of at most 10 points'

$$\therefore \quad n(A_3) = 33$$

$$P\text{ (At most 10 points)} = P(A_3)$$

$$= \frac{33}{36} = \frac{11}{12}$$

Examples on Probability

Example : 4

Three fair coins are tossed. Find the probability that at least two heads appear.

Solution :

Three fair coins are tossed.

$\therefore$ Sample space,

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

$$\therefore \quad n(S) = 8$$

let A = Event 'At least two heads appear'

$$= \{HHT, HTH, THH, HHH\}$$

$$\therefore \quad n(A) = 4$$

$\therefore$ P (At least two Heads appear)

$$= P(A)$$

$$= \frac{n(A)}{n(S)} = \frac{4}{8} = \frac{1}{2}$$

Example : 5

From 20 tickets marked 1 to 20, one ticket is drawn at random. Find the probability that it is marked with multiple of 3 or 5.

Solution :

$$\text{Sample space } S = \{1, 2, 3, \ldots\ldots, 20\}$$

$$n(S) = 20$$

One ticket is drawn from 20 tickets in $^{20}C_1$ ways

Let A = Event 'The ticket is marked with multiple of 3 or 5'.

$$= \{3, 6, 9, 12, 15, 18, 5, 10, 20\}$$

$$\therefore \quad n(A) = 9$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{^{9}C_1}{^{20}C_1} = \frac{9}{20}$$

Theorems on Probability

I) Theorem of Total Probability :

If A and B are any two events from a finite sample space 'S' then,

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

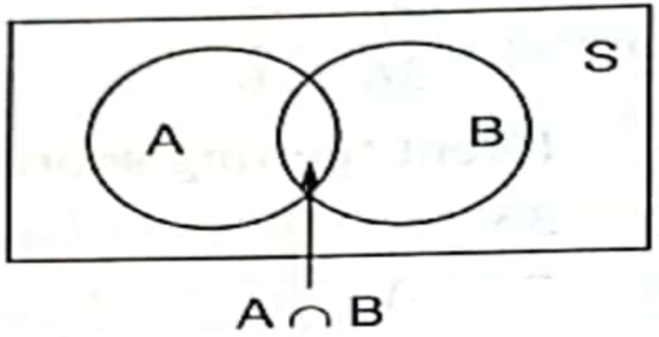


Fig. 3.4

- **Corrollary I :**

If A and B are mutually exclusive events then

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

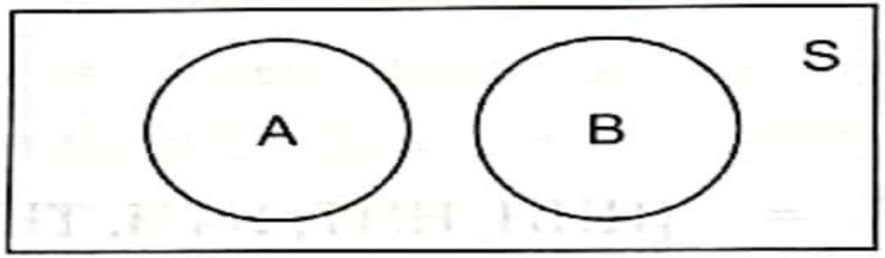


Fig. 3.5

Theorems on Probability

- **Corrollary II :**

  If $A_1, A_2, A_3, \ldots, An$ are mutually exclusive events then

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

- **Corrollary III :**

  If A and A′ are complementary events, then

  $$\begin{aligned} P(A \cup A^C) &= P(A) + P(A^C) - P(A \cap A') \\ &= P(A) + P(A^C) - 0 \\ &= P(A) + P(A^C) \end{aligned}$$

  Remark : $P(A) + P(A^C) = 1$

  or $P(A^C) = 1 - P(A)$

Theorems on Probability

II) Theorem of compound probability :

$$P(A \cap B) = P(A)\,P(B \mid A)$$

Where $B \mid A$ means occurrence of events, subject to the condition that A has already occurred.

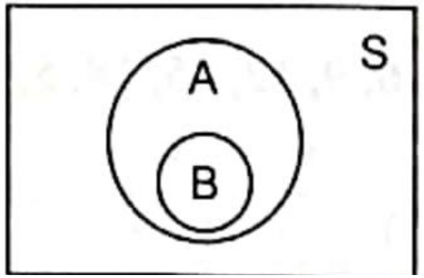


Fig. 3.6

Remark :

i) If $B \subset A$ then

$$P(A \cap B) = P(A)\,P(B \mid A)$$

ii) If $A \subset B$ then,

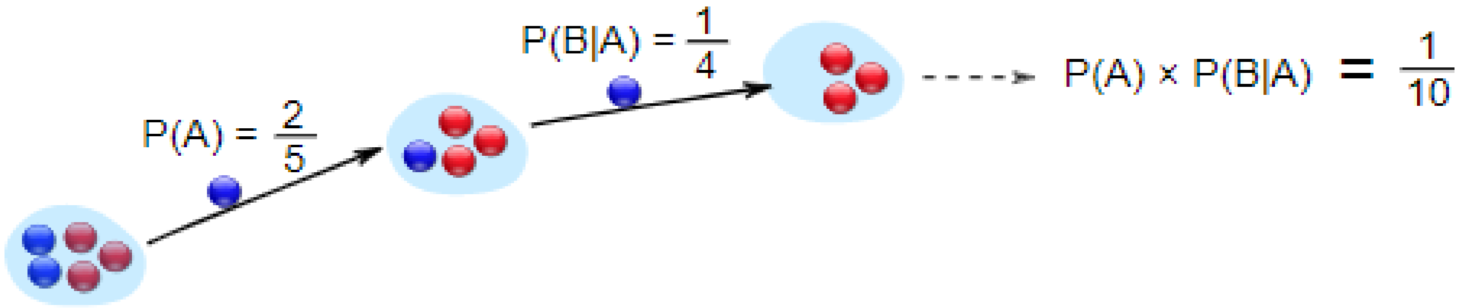
$$P(A \cap B) = P(B)\,P(A \mid B)$$

iii) If A and B are independent events then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$

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

Theorems on Probability

So the probability of getting **2 blue marbles** is:



And we write it as

"Probability Of" *"Given"*

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

Event A *Event B*

*"Probability of **event A and event B** equals*
*the probability of **event A** times the probability of **event B given event A**"*

III) Baye's theorem :

For any two events A & B.

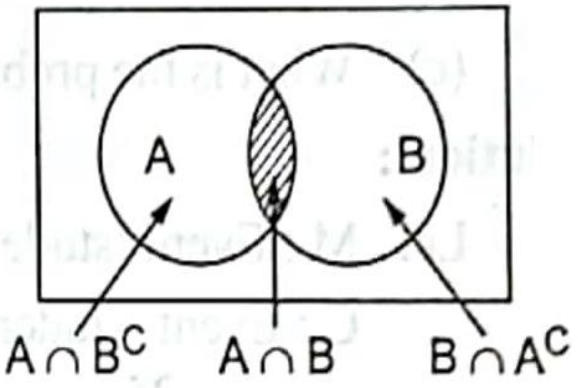


Fig. 3.7

$$P(A) = P(A \cap B^C) + P(A \cap B)$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{P(A|B).P(B)}{P(A \cap B^C) + P(A \cap B)}$$

$$= \frac{P(A|B).P(B)}{P(A|B^C)\,P(B^C) + P(A|B).P(B)}$$ is called Baye's formula.

In general,

$$P(A_K \mid A) = \frac{P(A_K)\,P(A \mid A_K)}{\Sigma P(A_K)\,P(A \mid A_K)}$$

$$K = 1, 2, \ldots.. n$$

where $A_1, A_2, \ldots\ldots A_n$ are mutually exclusive events whose union is S.

Examples

Example : 1

The probability that A hits a target is $\frac{1}{4}$, and the B hits the target is $\frac{2}{5}$. Both shoot at the target. Find the probability that at least one of them hits the target.

Solution : Here, $P(A) = \frac{1}{4}$, $P(B) = \frac{2}{5}$

Furthermore, we assume that A and B are independent events, that is the probability that A or B hits the target is not influenced by what the other does.

$$\therefore \quad P(A \cap B) = P(A)\,P(B)$$
$$= \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{10}$$

We want to find probability that at least one of them hits the target, i.e. that A or B (or both) hit the target

$$\therefore \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{1}{4} + \frac{2}{5} - \frac{1}{10} = \mathbf{\frac{11}{20}}$$

Examples

Example : 2

In a certain college town, 25 percent of the students failed Mathematics, 15 percent failed Chemistry and 10 percent failed both Mathematics and Chemistry. A student is selected at random.

(a) If the student failed Mathematics, what is the probability that he or she failed Chemistry?

(b) What is the probability that student failed Mathematics or Chemistry?

(c) What is the probability that the student failed neither Mathematics nor Chemistry?

Solution :

Let M : Event 'student failed Mathematics'

C : Event 'student failed Chemistry'

$\therefore \quad P(M) = \dfrac{25}{100} = 0.25, \; P(C) = \dfrac{15}{100} = 0.15, \; P(M \cap C) = 0.10$

(a) The probability that student failed Chemistry, given that he or she failed Mathematics is P(C/M)

we know that,

$$P(M \cap C) = P(C \mid M) . P(M)$$

$$\therefore \quad P(C \mid M) = \frac{P(M \cap C)}{P(M)} = \frac{0.10}{0.25} = \frac{2}{5}$$

Examples

(b) P (Student failed Mathematics or Chemistry)

$$= P(M \cup C)$$

$$= P(M) + P(C) - P(M \cap C)$$

$$= 0.25 + 0.15 - 0.10 = 0.30$$

(c) P (Student failed neither Mathematics nor Chemistry)

$$= P(M \cup C)^C$$

$$= 1 - P(M \cup C)$$

$$= 1 - 0.30 = \mathbf{0.70}$$

PREREQUISITES

$^{n}C_{r}$ = Selection of r objects from n objects (without repetition)

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

Examples

Example : 3

A class has 12 boys and 4 girls. Suppose three students are selected at random from the class. Find the probability that they are all boys.

Solution :

Total students = 12 + 4 = 16

Three students can be selected out of 16 students in ${}^{16}C_3$ ways.

$\therefore$ Total number of ways = ${}^{16}C_3 = 560$

$${}^{16}C_3 = \frac{16!}{3!(16-3)!}$$

Three Boys can be selected out of 12 boys in ${}^{12}C_3$ ways.

$\therefore$ Number of favourable ways = ${}^{12}C_3 = 220$

$\therefore$ P (All three selected are boys) $= \dfrac{220}{560} = \dfrac{22}{56} = \dfrac{11}{28}$

Examples

Example : 4

Box A contains 5 red marbles and 3 blue marbles, and Box B contains 3 red and 2 blue. A marble is drawn at random from each box.

(a) Find the probability that both marbles are red.

(b) Find the probability that one is red and one is blue.

Solution :

Let

(a) P(A) = Probability of choosing red marble from box A $= \frac{5}{8}$

P(B) = Probability of choosing red marble from box B $= \frac{3}{5}$

Since both events are independent.

$\therefore$ Probability that both marbles are red

$$= P(A \cap B)$$
$$= P(A) \cdot P(B)$$
$$= \frac{5}{8} \cdot \frac{3}{5} = \mathbf{\frac{3}{8}}$$

(b) The probability of choosing a red marble from A and a blue marble from B is

$$P_1 = \frac{5}{8} \cdot \frac{2}{5} = \frac{1}{4}$$

The probability of choosing a blue marble from A and a red marble from B is

$$P_2 = \frac{3}{8} \cdot \frac{3}{5} = \frac{9}{40}$$

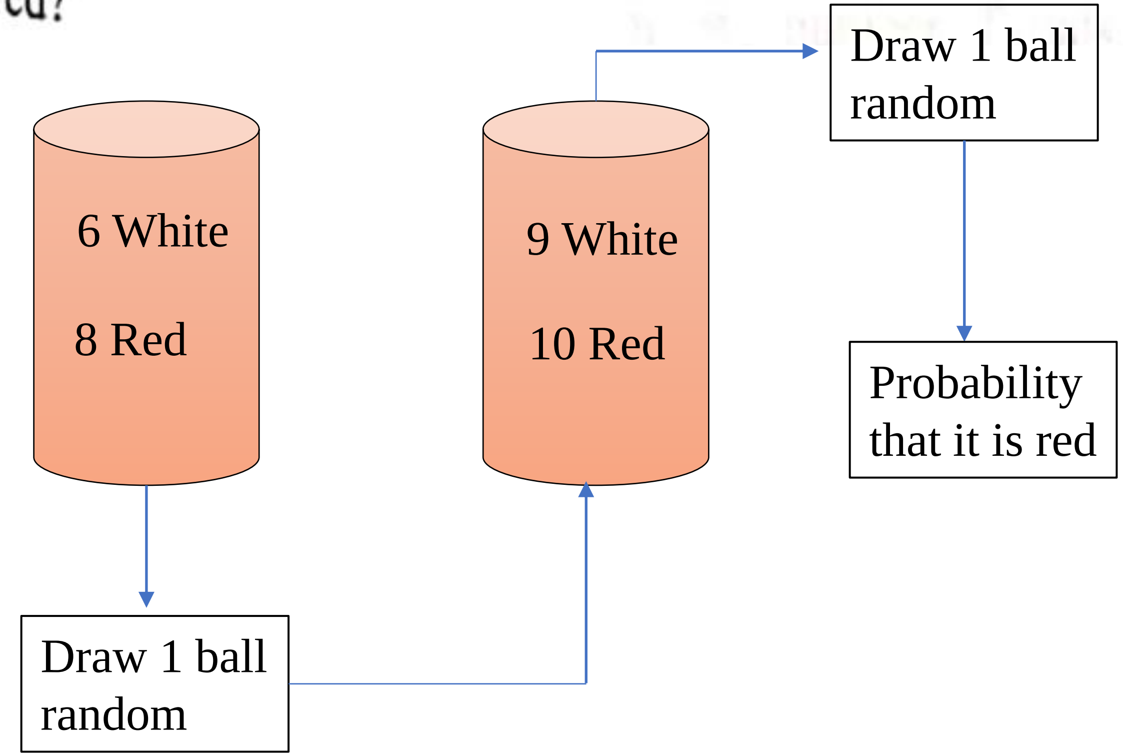
Hence, P = Probability that one is red and one is blue

$$P = P_1 + P_2$$
$$P = \frac{1}{4} + \frac{9}{40} = \mathbf{\frac{19}{40}}$$

Examples

Example : 5

An urn contains 6 white and 8 red balls. Second urn contains 9 white and 10 red balls. One ball is drawn at random from the first urn and put into the second urn without noticing its colour. A ball is then drawn at random from the second urn. What is the probability that it is red?



Solution :

1) The White ball is transferred from the first urn to the second and then red ball is drawn from it

2) The red ball is transferred from the first urn to the second and then red ball is drawn from it

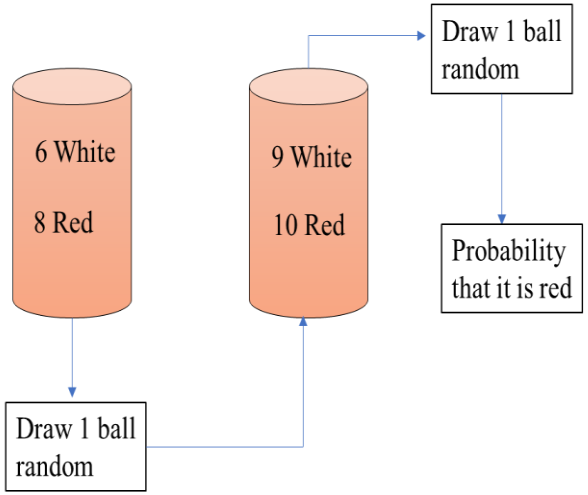
Examples

Let,

A = Event 'Transferring white ball from the first urn'

B = Event 'Transferring a red ball from the first urn'

C = Event 'drawing a red ball from the second urn'



$$P(A) = \frac{6}{14} = \frac{3}{7}$$

$$P(B) = \frac{8}{14} = \frac{4}{7}$$

$$P(C \mid A) = \frac{10}{20} = \frac{1}{2} \quad (\because \text{ second urn contain 10 white \& 10 red balls})$$

$$P(C \mid B) = \frac{11}{20} \quad (\because \text{ second urn contain 9 white \& 11 red balls})$$

$$\begin{aligned} P_1 = P(i) &= P(A \cap C) \\ &= P(A)\, P(C \mid A) \\ &= \frac{3}{7} \times \frac{1}{2} = \frac{3}{14} \end{aligned}$$

$$\begin{aligned} P_2 = P(ii) &= P(B \cap C) \\ &= P(B) \cdot P(C \mid B) \\ &= \frac{4}{7} \times \frac{11}{20} = \frac{11}{35} \end{aligned}$$

Examples

Required probability,

$$P = P_1 + P_2$$

$$= \frac{3}{14} + \frac{11}{35}$$

$$= \frac{259}{490}$$

Examples

Ex. 6 : *Two cards are drawn from a well shuffled pack of 52 cards. Find the probability that they are both kings if*

(i) the first card drawn is replaced, (ii) first card drawn is not replaced.

Sol. : (i) First card drawn is replaced : Let A_1 be event of king on first draw and A_2 be the event of king on second draw. $A_1 \cap A_2$ is the event of king at both draws.

If first card drawn is replaced before the second draw then A_1, A_2 are independent.

$$\therefore \quad P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

$$P(A_1) = \frac{4}{52} = \frac{1}{13} \text{ (as there are four kings)}$$

$$P(A_2) = \frac{4}{52} = \frac{1}{13}$$

$$\therefore \quad P(A_1 \cap A_2) = \frac{1}{13} \cdot \frac{1}{13} = \frac{1}{169}$$

Examples

(ii) First card drawn is not replaced : In this case, the two events are dependent.

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1)$$

$$\therefore \quad P(A_1) = \frac{4}{52} = \frac{1}{13}$$

If first draw is king then for second draw, 3 kings are left to be chosen out of 51 cards.

$$P(A_2 | A_1) = \frac{3}{51} = \frac{1}{17}$$

$$\therefore \quad P(A_1 \cap A_2) = \frac{1}{13} \cdot \frac{1}{17} = \frac{1}{221}$$

Random Variable (X) :

It is a real valued function defined over the sample space of an experiment i.e. $X : S \to R$

1) Binomial Probability Distribution:

Consider an experiment with only two outcomes, one is called success and other is called as failure. Independent repeated trials of such experiment is called as Bernoulli trails.

Let p be probability of success and $q = 1 - p$ is probability of failure

A Binomial experiment consist of fixed number of trails r

Probability of r success in n trails using Binomial distribution is given by

$$\boxed{p(r) = {}^nC_r\, p^r q^{n-r}} \quad where,\ p + q = 1$$

Probability Distribution-Binomial Probability Distribution

Mean & Variance of Binomial Distribution :

1) $\text{mean} = \mu = np$

2) $\text{variance} = v = \sigma^2 = npq$

3) $\text{standard deviation} = \sigma = \sqrt{npq}$

4) Sum of all probabilities is one

Q.1 A fair coin is tossed 6 times. Find a probability of getting

a) exactly two heads
b) at least four heads
c) not heads
d) one or more head

Solution :

Let p = probability of getting head

$$\boxed{\Rightarrow p = \frac{1}{2}} \qquad \Rightarrow q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2} \qquad \boxed{\Rightarrow q = \frac{1}{2}}$$

By Binomial distribution $p(r) = {}^{n}C_r\, p^r q^{n-r}\ , n = 6$

Case (a) : exactly two heads $p(r=2) = C_2\, p^2 q^{6-2} = {}^{6}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = 0.2344$

Case (b) : at least four heads

$$p(r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

$$, n = 6$$

$$p(r \ge 4) = p(r = 4) + p(r = 5) + p(r = 6)$$

$$= {}^{6}C_{4}\, p^{4} q^{6-4} + {}^{6}C_{5}\, p^{5} q^{6-5} + {}^{6}C_{6}\, p^{6} q^{6-6}$$

$$= {}^{6}C_{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{2} + {}^{6}C_{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{1} + {}^{6}C_{6}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{0}$$

$$= 0.3438$$

Case (c) : not head

$$p(r = 0) = {}^{6}C_{0}\, p^{0} q^{6-0}$$

$$= {}^{6}C_{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{6}$$

$$= 0.0156$$

Case (c) : one or more head

$$p(r \geq 1) = p(r=1) + p(r=2) + p(r=3) + p(r=4) + p(r=5) + p(r=6)$$

We know that sum of all probabilities is 1

$$i.e.\quad \mathrm{p}(r=0) + p(r=1) + p(r=2) + p(r=3) + p(r=4) + p(r=5) + p(r=6) = 1$$

$$\Rightarrow p(r \geq 1) = 1 - p(r=0)$$

$$= 1 - {}^{6}C_{0}\, p^{0} q^{6-0}$$

$$= 1 - {}^{6}C_{0} \left(\frac{1}{2}\right)^{0} \left(\frac{1}{2}\right)^{6}$$

$$= 0.9844$$

Q.2 On an average a box containing 10 articles is likely to have 2 defectives. If we consider a consignment of 100 boxes, how many of them are expected to have three or less defective

Solution :

Let p = probability of defective articles in box

$$p = \frac{2}{10} = \frac{1}{5} \quad \boxed{\Rightarrow p = \frac{1}{5}} \qquad \& \; q = 1 - p = 1 - \frac{1}{5} = \frac{4}{5} \quad \boxed{\Rightarrow q = \frac{4}{5}}$$

By Binomial distribution $p(r) = {}^nC_r\, p^r q^{n-r} \quad , n = 10$

Probability of one or more to be defective from box

$$p(r \le 3) = p(r=0) + p(r=1) + p(r=2) + p(r=3)$$

$$p(r \le 3) = p(r=0) + p(r=1) + p(r=2) + p(r=3)$$

$$= {}^{10}C_0 p^0 q^{10-0} + {}^{10}C_1 p^1 q^{10-1} + {}^{10}C_2 p^2 q^{10-2} + {}^{10}C_3 p^3 q^{10-3}$$

$$= {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10} + {}^{10}C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^9 + {}^{10}C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^8 + {}^{10}C_3 \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^7$$

$$= 0.8791$$

Number of defective articles from consignment of 100 articles is

$$= 100 \times 0.8791$$

$$= 87.91$$

Q.3 The probability of a man hitting target is $\frac{1}{3}$. If he fires 5 times, what is the probability of his hitting the target at least twice

Solution :

Let p = probability of man hitting the target

$$\boxed{p=\frac{1}{3}} \qquad \& \; q=1-p=1-\frac{1}{3}=\frac{2}{3} \qquad \boxed{\Rightarrow q=\frac{2}{3}}$$

By Binomial distribution $p(r)={}^{n}C_{r}\,p^{r}q^{n-r} \quad ,n=5$

Probability of man hitting at least target at least twice

$$p(r\geq 2)=p(r=2)+p(r=3)+p(r=4)+p(r=5)$$

Probability Distribution-Binomial Probability Distribution

We know that sum of all probabilities is 1

$$i.e.\quad p(r=0)+p(r=1)+p(r=2)+p(r=3)+p(r=4)+p(r=5)=1$$

$$\boxed{\Rightarrow q=\frac{2}{3}} \quad \boxed{p=\frac{1}{3}}$$

$$p(r\geq 2)=p(r=2)+p(r=3)+p(r=4)+p(r=5)$$

$$=1-\left[p(r=0)+p(r=1)\right]$$

$$=1-\left[{}^5C_0 p^0 q^{5-0}+{}^5C_1 p^1 q^{5-1}\right]$$

$$=1-\left[{}^5C_0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5+{}^5C_1\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^4\right]$$

$$=0.539$$

Q.4 20% of bolts produced by a machine is defective. Determine the probability that out of 4 bolts chosen at random (i) one is defective (ii) zero is defective (iii) at most two bolts are defective

Solution :

Let p = probability of defective bolt

$$p = \frac{20}{100} = \frac{1}{5} \quad \boxed{\Rightarrow p = \frac{1}{5}} \qquad \& \; q = 1 - p = 1 - \frac{1}{5} = \frac{4}{5} \quad \boxed{\Rightarrow q = \frac{4}{5}}$$

By Binomial distribution $p(r) = {}^{n}C_{r}\, p^{r} q^{n-r} \quad , n = 4$

Case (i) : one is defective

$$p(r=1) = {}^{4}C_{1}\, p^{1} q^{4-1} = {}^{4}C_{1} \left(\frac{1}{5}\right)^{1} \left(\frac{4}{5}\right)^{3} = 0.4096$$

Probability Distribution-Binomial Probability Distribution

Case (ii) : zero are defective

$$p(r=0) = {}^4C_0\, p^0 q^{4-0} = {}^4C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^4 = 0.4096$$

Case (ii) : at most 2 are defective

$$p(r \le 2) = p(r=0) + p(r=1) + p(r=2)$$

$$= {}^4C_0\, p^0 q^{4-0} + {}^4C_1\, p^1 q^{4-1} + {}^4C_2\, p^2 q^{4-2}$$

$$= {}^4C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^4 + {}^4C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^3 + {}^4C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^2$$

$$= 0.9728$$

Q.5 Out of 2000 families with 4 children each, how many would you expect to have (i) at least one boy (ii) 2 boys (iii) 1 or 2 girls (iv) no girls.

Solution :

Let p = probability of having boy for one family

$$\boxed{p=\frac{1}{2}} \qquad \& \; q=1-p=1-\frac{1}{2}=\frac{1}{2} \qquad \boxed{\Rightarrow q=\frac{1}{2}}$$

By Binomial distribution $\quad p(r)={}^{n}C_{r}\,p^{r}q^{n-r} \quad ,n=4$

Case (i) : at least one boy

$$p(r\geq 1)=p(r=1)+p(r=2)+p(r=3)+p(r=4)$$

$$=1-p(r=0)$$

$$p(r \geq 1) = 1 - p(r = 0)$$

$$= 1 - {}^4C_0 p^0 q^{4-0}$$

$$= 1 - {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4$$

$$= 0.9375$$

2000 families having at least one boy

$$= 2000 \times 0.9375$$

$$= 1875$$

Case (ii) : 2 boys

$$p(r = 2) = {}^4C_2 p^2 q^{4-2} = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = 0.3750$$

Probability Distribution-Binomial Probability Distribution

2000 families with 2 boys

$$= 2000 \times 0.3750$$

$$= 750$$

Let p = probability of having boy for one family

Case (iii) : 1 or 2 girl

Probability of having 1 or 2 girl $= p(3\ or\ 2\ boys)$

$$= p(r=3) + p(r=2)$$

$$= {}^4C_3 p^3 q^{4-3} + {}^4C_2 p^2 q^{4-2}$$

$$= {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1 + {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$$

$$= 0.6250$$

Probability Distribution-Binomial Probability Distribution

2000 families with 1 or 2 girls

$$= 2000 \times 0.6250$$

$$= 1250$$

Case (iv) : No girl

Probability of 4 are boys $= p(r = 4)$

$$= {}^4C_4 p^4 q^{4-4}$$

$$= {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0$$

$$= 0.0625$$

2000 families with "No girls" $= 2000 \times 0.0625 = 125$

Q.6 The mean and variance of Binomial distribution are 6 and 2 respectively. Find $p(r \geq 1)$.

Solution :

We know that

$$1)\ \text{mean} = \mu = np = 4 \quad \boxed{\Rightarrow np = 6.........(1)}$$

$$2)\ \text{variance} = v = \sigma^2 = npq = 2 \quad \boxed{\Rightarrow npq = 2.........(2)}$$

$\therefore$ we get

$$q = \frac{2}{6} = \frac{1}{3} \quad \boxed{\Rightarrow q = \frac{1}{3}} \quad \Rightarrow p = 1 - \frac{1}{3} = \frac{2}{3} \quad \boxed{\Rightarrow p = \frac{2}{3}}$$

From equation (1) & (2)

From equation (1) $\boxed{\Rightarrow n = 9}$

$$\therefore\ p(r \geq 1) = 1 - p(r = 0)$$

$$= 1 - {}^{9}C_{0}\, p^{0} q^{9-0}$$

$$= 1 - {}^{9}C_{0} \left(\frac{2}{3}\right)^{0} \left(\frac{1}{3}\right)^{9}$$

$$= 0.9999$$

2) Poisson Distribution :

When p be probability of success is very small and n the number of trials is very large and mean np is finite then we get another distribution called Poisson Distribution.

It is considered as limiting case of Binomial distribution with $n \to \infty, p \to 0$ and np remaining finite.

Let $z = np$

The probability of r success by Poisson Distribution is given by

$$\boxed{p(r) = \frac{z^r e^{-z}}{r!}}$$

Remarks :

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

1) Mean $= np = z$

2) Variance $= npq = z$ ($\lim q = 1$ as $p \to 0$)

3) Standard Deviation $= \sqrt{z}$

$$\boxed{Mean = E(X) = \sum_{r=0}^{\infty} r p(r)}$$

$$= \sum_{r=0}^{\infty} r \frac{e^{-z} z^r}{r!} = e^{-z} \sum_{r=0}^{\infty} \frac{z^r}{(r-1)!} = e^{-z} z \sum_{r=0}^{\infty} \frac{z^{r-1}}{(r-1)!}$$

$$= e^{-z} z e^{z}$$

$$= z$$

Probability Distribution-Poisson Distribution

$$E\left(X^2\right)=\sum_{r=0}^{\infty} r^2 p(r) \quad = z^2 + z$$

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

$$\boxed{vax\left(X\right)=E\left(X^2\right)-E\left(X\right)^2}$$

$$= z^2 + z - z^2$$

$$= z$$

Q.1 In a certain factory turning out razor blades, there is s small chance of $\frac{1}{500}$ for any blade to be defective. The blades are supplied in a packet of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective and two defective blades, in a consignment of 10,000 packets.

Solution :

Let p = probability of defective blade

$$p = \frac{1}{500} = 0.002 \qquad \& \ q = 1 - p \quad \Rightarrow q = 1 - 0.002 = 0.998$$

$$\boxed{\therefore \ p = 0.002, q = 0.998}$$

As p is very small, therefore by Poisson Distribution $\boxed{p(r) = \frac{e^{-z} z^{r}}{r!}}, n = 10$

Probability Distribution-Poisson Distribution

$$p(r) = \frac{z^r e^{-z}}{r!}$$

where

$\boxed{z = np}$ $\quad \Rightarrow z = 10 \times 0.002 \quad$ $\boxed{\Rightarrow z = 0.02}$

Case I : No defective blade

$$p(r=0) = \frac{e^{-z} z^0}{0!} = \frac{e^{-0.02}}{0!} = 0.9802$$

$\therefore$ Number of packets containing no defective blades in a consignment of 10,000 packets

$$= 10,000 \times 0.9802$$

$$= 9801.9867 \approx 9802$$

Probability Distribution-Poisson Distribution

$$p(r)=\frac{z^r e^{-z}}{r!}$$

Case II : Two defective blade

$$p(r=2)=\frac{e^{-z}z^2}{2!}=\frac{e^{-0.02}(0.02)^2}{2!}=0.0002$$

$\therefore$ Number of packets containing two defective blades in a consignment of 10,000 packets

$$=10,000\times 0.0002$$

$$=1.9604 \approx 2$$

Q.2 A manufacture of cotter pins knows that 2% of his product is defective. If he sells cotter pins and guarantees that not more than 5 pins will be defective in a box, find the approximate probability that a box will fail to meet guaranteed quality.

Solution :

Let p = probability of cotter pin to be defective

$$p = \frac{2}{100} = 0.02 \qquad \& \; q = 1 - p \qquad \Rightarrow q = 1 - 0.02 = 0.98$$

$$\boxed{\therefore p = 0.02, q = 0.98}$$

As p is very small, therefore by Poisson Distribution $\boxed{p(r) = \frac{e^{-z} z^{r}}{r!}}$ $, n = 100$

where

$$\boxed{z = np} \qquad \Rightarrow z = 100 \times 0.02 \qquad \boxed{\Rightarrow z = 2}$$

Probability Distribution-Poisson Distribution

$$p(r)=\frac{z^r e^{-z}}{r!}$$

$$\Rightarrow z=2$$

$\therefore$ Probability of not more than 5 cotter pin to be defective

$$p(r\le 5)=p(r=0)+p(r=1)+p(r=2)+p(r=3)+p(r=4)+p(r=5)$$

$$=\frac{e^{-2}2^0}{0!}+\frac{e^{-2}2^1}{1!}+\frac{e^{-2}2^2}{2!}+\frac{e^{-2}2^3}{3!}+\frac{e^{-2}2^4}{4!}+\frac{e^{-2}2^5}{5!}$$

$$=0.9834$$

$\therefore$ Probability that a box will fail to meet guaranteed quality

$$p(r>5)=1-p(r\le 5)=1-0.9834\ \boxed{=0.0166}$$

Q.3 The average number of misprints per page of book is 1.5 . Assuming the distribution of number of misprints to be poisson, find

a) The probability that a particular book is free from misprints
b) Number of pages containing more than one misprint if the book contain 900 pages

Solution :

Since, The average number of misprints per page of book is 1.5

$$\Rightarrow z = 1.5 = np$$

Where, p = probability of misprint per page

By Poisson Distribution

$$p(r) = \frac{e^{-z} z^r}{r!}$$

Probability Distribution-Poisson Distribution

Case I : Probability that a particular book is free from misprints

$$p(r=0)=\frac{e^{-1.5}(1.5)^{0}}{0!}=0.2231$$

Case II : Probability that a particular book contain more than one misprint per page

$$p(r>1)=1-p(r\le 1)$$

$$=1-\left[p(r=0)+p(r=1)\right]$$

$$=1-\left[\frac{e^{-1.5}(1.5)^{0}}{0!}+\frac{e^{-1.5}(1.5)^{1}}{1!}\right]$$

$$=1-0.5578=0.4422$$

Number of pages containing more than one misprint in a book containing 900 pages = $900 \times 0.4422 = 397.98 \sim 398$

Q.4 Number of road accidents on a high way during a month follows a Poisson Distribution with mean 5,find the probability that in certain month number of accidents on the highway will be (i) Less than 3 (ii) Between 3 & 5 (iii) More than 3.

Solution :

We have

$$z = 5 = np$$

Where, p = probability of accident in one month

By Poisson Distribution

$$p(r) = \frac{e^{-z} z^{r}}{r!}$$

Case I : Probability that in certain month number of accidents less than 3

$$p(r<3) = p(r=0) + p(r=1) + p(r=2)$$

$$= \frac{e^{-5}\ 5^{0}}{0!} + \frac{e^{-5}\ 5^{1}}{1!} + \frac{e^{-5}\ 5^{2}}{2!}$$

$$= 0.1247$$

Case II : Probability that in certain month number of accidents between 3 & 5

$$p(3 \le r \le 5) = p(r=3) + p(r=4) + p(r=5)$$

$$= \frac{e^{-5}\ 5^{3}}{3!} + \frac{e^{-5}\ 5^{4}}{4!} + \frac{e^{-5}\ 5^{5}}{5!}$$

$$= 0.4913$$

Case III : Probability that in certain month number of accidents more than 3

$$p(r>3)=1-p(r\le 3)$$

$$=1-\left[p(r=0)+p(r=1)+p(r=2)+p(r=3)\right]$$

$$=1-\left[\frac{e^{-5}\ 5^{0}}{0!}+\frac{e^{-5}\ 5^{1}}{1!}+\frac{e^{-5}\ 5^{2}}{2!}+\frac{e^{-5}\ 5^{3}}{3!}\right]$$

$$=1-0.2650$$

$$=0.7350$$

Q.5 The accidents per shift in a factory are given by the table

Accidents x per shift	0	1	2	3	4	5
Frequency f	142	158	67	27	5	1

Fit a Poisson Distribution to the above table and calculate theoretical frequencies.

Solution :

$$\text{Mean} = z = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

$$= \frac{(0\times142)+(1\times158)+(2\times67)+(3\times27)+(4\times5)+(5\times1)}{142+158+67+27+5+1}$$

$$= 0.9950$$

Total frequency = 400

Probability Distribution-Poisson Distribution

By Poisson Distribution $p(r) = \dfrac{e^{-z} z^{r}}{r!}$

Theoretical (expected) frequencies= $Total\ frequencies \times Probability$

x	Observed frequency (Given)	Theoretical frequency
0	142	$= 400 \times p(r=0) = 400 \times \dfrac{e^{-0.995}(0.995)^{0}}{0!} = 147.88 \approx 148$
1	158	$= 400 \times p(r=1) = 400 \times \dfrac{e^{-0.995}(0.995)^{1}}{1!} = 147.1499 \approx 147$
2	67	$= 400 \times p(r=2) = 400 \times \dfrac{e^{-0.995}(0.995)^{2}}{2!} = 73.2071 \approx 73$
3	27	$= 400 \times p(r=3) = 400 \times \dfrac{e^{-0.995}(0.995)^{3}}{3!} = 24.2804 \approx 24$
4	5	$= 400 \times p(r=4) = 400 \times \dfrac{e^{-0.995}(0.995)^{4}}{4!} = 6.0397 \approx 6$
5	1	$= 400 \times p(r=5) = 400 \times \dfrac{e^{-0.995}(0.995)^{5}}{5!} = 1.2019 \approx 1$

Q.6 In a Poisson distribution if $p(r=1)=2p(r=2)$, find $p(r=3)$.

Solution :

By Poisson Distribution

$$p(r)=\frac{e^{-z}z^{r}}{r!}$$

$$p(r=1)=2p(r=2)$$

$$\Rightarrow \frac{e^{-z}z^{1}}{1!}=2\frac{e^{-z}z^{2}}{2!}$$

$$\Rightarrow z=z^{2}$$

$$\Rightarrow z-z^{2}=0$$

$$\Rightarrow z(1-z)=0 \quad \boxed{\Rightarrow z=1}\left(\because z\neq 0\right)$$

$$\therefore p(r=3)=\frac{e^{-z}z^{3}}{3!}$$

$$=0.0613$$

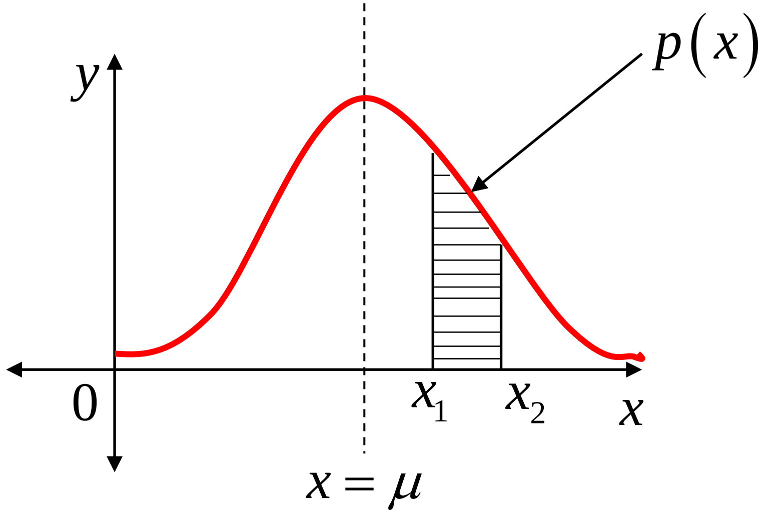
3) Normal Distribution :

Normal distribution is obtained as limiting form of Binomial distribution when n the number of trials is very large and neither p nor q is very small.

Most of the modern statistical methods have been based on this distribution.

The normal distribution curve is given by

$$\boxed{p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$$



Remarks :

1) $\int_{-\infty}^{\infty} y\,dx = 1$

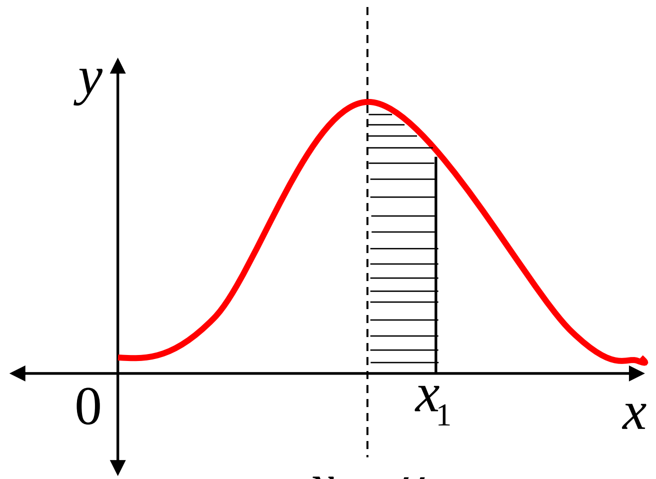
2) $p(x_1 < x < x_2)$ = Area under the curve from $x = x_1$ to $x = x_2$

3) $p(x_1) = p(\mu < x < x_1)$ = Area under the curve from $x = \mu$ to $x = x_2$

Transformation :

We know that

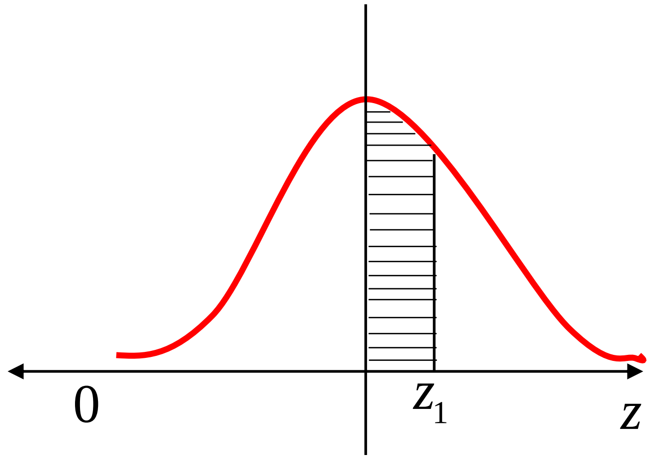
$$p(x_1) = p(\mu < x < x_1) = \int_{\mu}^{x_1} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$



Put $z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{1}{\sigma} dx$

x	μ	x_1
z	0	$z_1 = \frac{x-\mu}{\sigma}$

$$p(z_1) = p(0 < z < z_1) = \int_{0}^{z_1} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$$

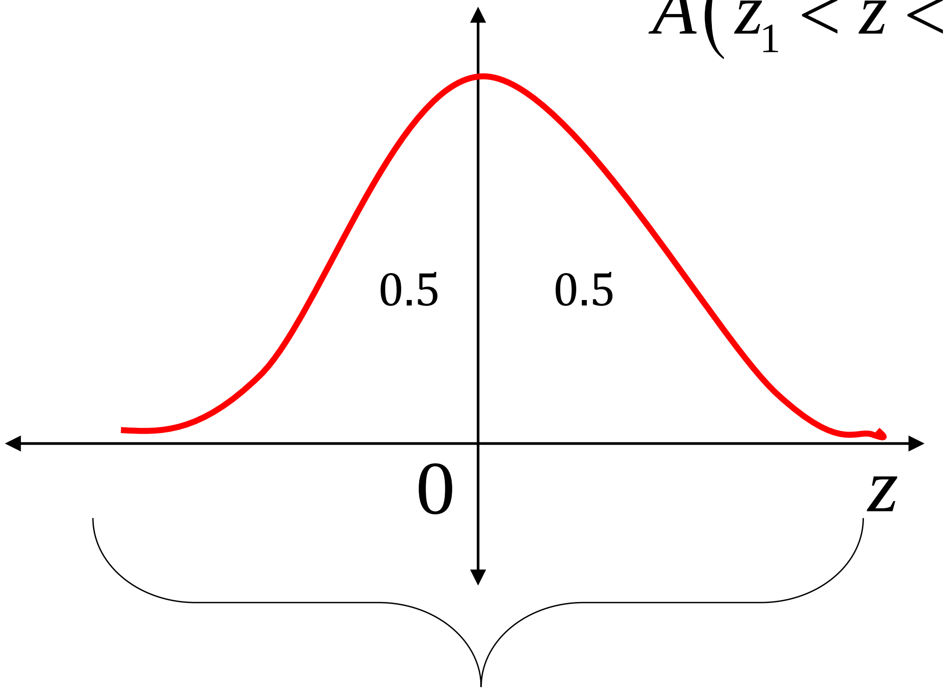


Probability Distribution-Normal Distribution

Area of Normal Curve :

$$A(z_1) = \text{Area from 0 to } z_1$$

$$A(z_1 < z < z_2) = \text{Area from } z_1 \text{ to } z_2$$



Total area = 1

Q.1 Assuming that the distance of 1000 brass plugs taken consecutively from machine form a normal distribution with mean 0.7515 cm and standard deviation 0.0020 cm. How many of the plugs are likely to be approved if the acceptable diameter is 0.752±0.004 cm
[Given : Area =0.4878 for $z = 2.25$,Area = 0.4599 for $z = 1.75$]

Solution :

Given that

$$\mu = 0.7515 \quad , \sigma = 0.0020$$

Let $x_1 = 0.752 + 0.004 = 0.756$

& $x_2 = 0.752 - 0.004 = 0.748$

Claim : $p\left(x_2 < x < x_1\right)$

Probability Distribution-Normal Distribution

Using transformation

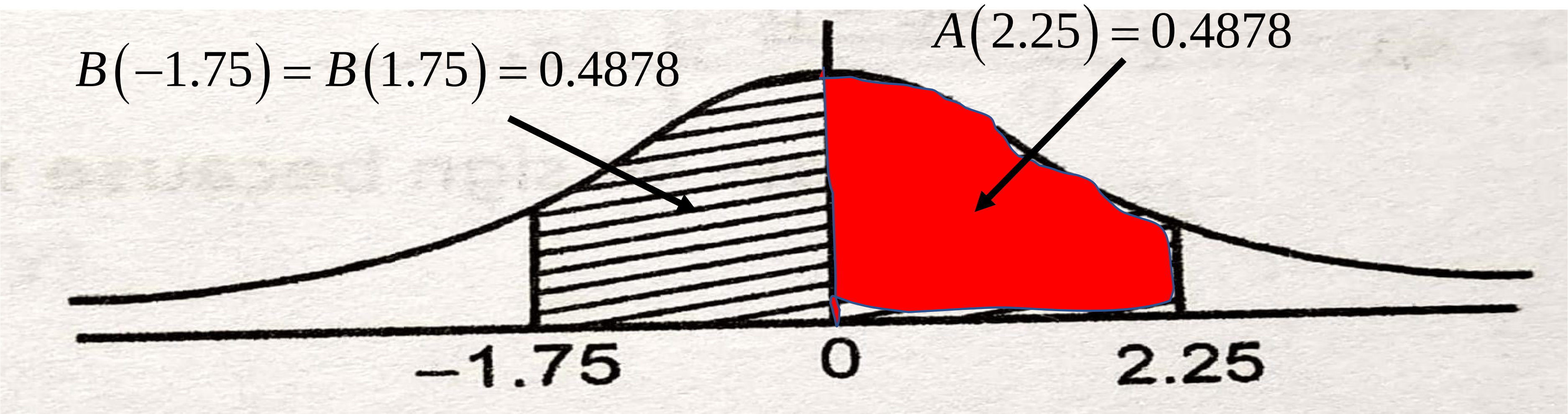
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{0.756 - 0.7515}{0.0020} = 2.25$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{0.748 - 0.7515}{0.0020} = -1.75$$

$$\mu = 0.7515$$
$$\sigma = 0.0020$$
$$x_1 = 0.756$$
$$\& \; x_2 = 0.748$$

[Given : Area =0.4878 for $z = 2.25$,Area = 0.4599 for $z = 1.75$]

$$\therefore p\left(x_2 < x < x_1\right) = p\left(z_2 < z < z_1\right) = p\left(-1.75 < z < 2.25\right) = A\left(2.25\right) + B\left(-1.75\right)$$



$$p(z_1 < z < z_2) = A(2.25) + B(-1.75)$$
$$= 0.4878 + 0.4599$$
$$= 0.9477$$

Number of plugs likely to be approved $= 1000 \times 0.9477 = 947.7 \sim \mathbf{948}$

Q.2 In a certain examination test, 2000 students appeared in a subject of statistics. Average marks obtained were 50% with standard deviation 5%. How many students do you expect to obtain more than 60% of marks, supposing that marks are distributed normally?

Given : Area $=0.4772$ for $z = 2$ or $A(2) = 0.4772$

Solution :

Given that

$$\mu = \frac{50}{100} = 0.5 \quad , \sigma = \frac{5}{100} = 0.05$$

Let $\quad x_1 = \frac{60}{100} = 0.6$

Claim : $p(x > x_1) = p(x > 0.6)$

Probability Distribution-Normal Distribution

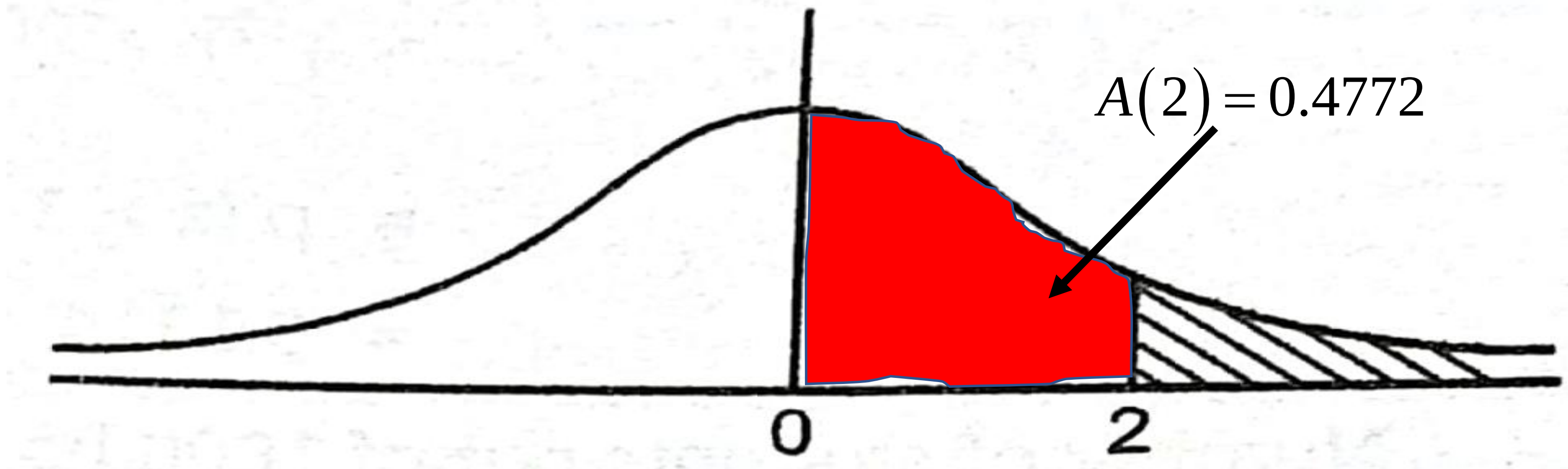
$$\mu = 0.5$$
$$\sigma = 0.05$$
$$x_1 = 0.6$$

Using transformation

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{0.6 - 0.5}{0.05} = 2$$

Given : Area =0.4772 for $z = 2$ or $A(2) = 0.4772$

$$\therefore p(x > x_1 = 0.6) = p(z > z_1 = 2)$$
$$= p(z > 2)$$
$$= 0.5 - A(2) = 0.5 - 0.4772 \boxed{= 0.0228}$$



Probability Distribution-Normal Distribution

Number of students expected to get more than 60% marks

$$= 2000 \times 0.0228 = 45.6 \sim \mathbf{46}$$

Q.3 In certain city 4000 tube lights are installed. If the lamps have average life of 1500 burning hours with standard deviation 100 hours. Assuming normal distribution

a) How many lamps will fail in first 1400 hours
b) How many lamps will last beyond 1600 hours

Given : $A(1) = 0.3413$

Solution :

Given that

$$\mu = 1500 \quad , \sigma = 100$$

Let $x_1 = 1400$ & $x_2 = 1600$

Claim : $p\left(x = x_1 = 1400\right)$ & $p\left(x > x_2 = 1600\right)$

Probability Distribution-Normal Distribution

Using transformation

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1400 - 1500}{100} = -1$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{1600 - 1500}{100} = 1$$

$$\mu = 1500$$

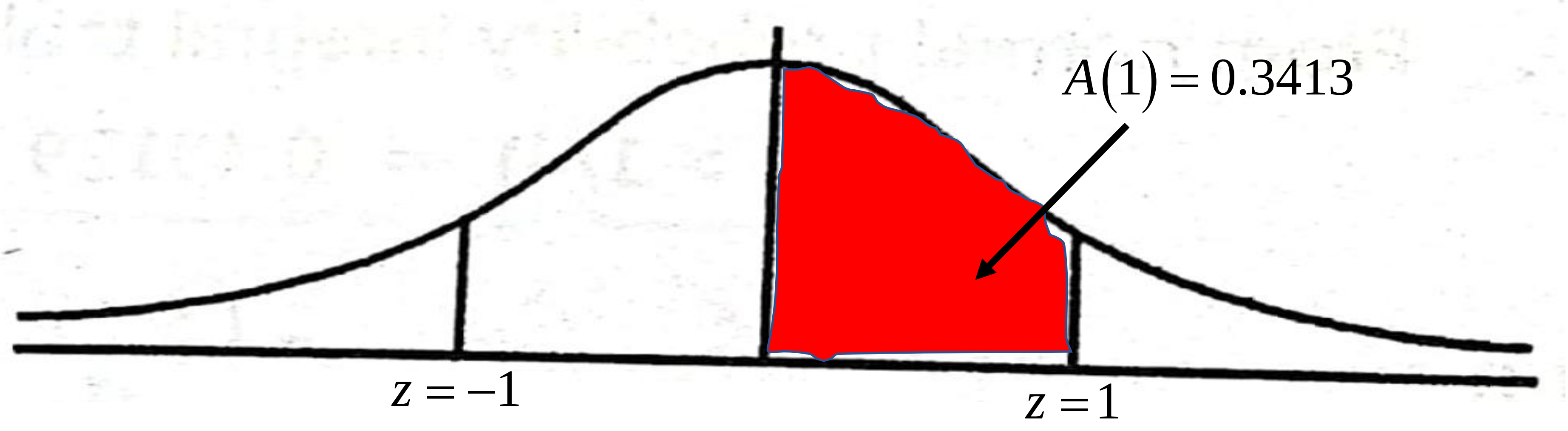
$$\sigma = 100$$

$$x_1 = -1$$

$$\& \ x_2 = 1$$

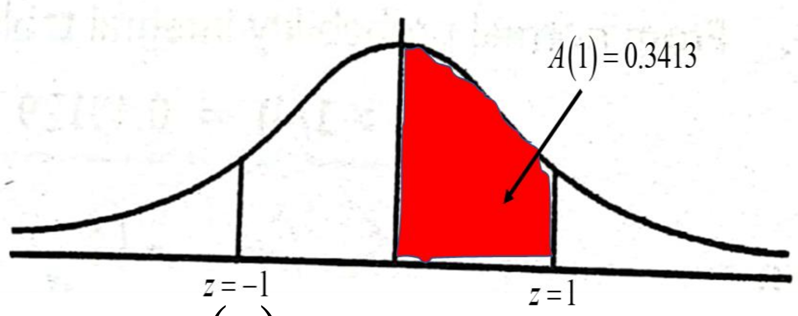
Given : $A(1) = 0.3413$

$$\therefore p\left(x = x_1 = 1400\right) = p\left(z = z_1 = -1\right) = A\left(-1\right) = A\left(1\right) = 0.3413$$



Probability Distribution-Normal Distribution

Number of tubes fail in first 1400 hours $= 4000 \times 0.3413 = 1365.2 \sim \mathbf{1365}$



$$\& \, p(x > x_2 = 1600) = p(z > z_1 = 1)$$

$$= A(1)$$

$$= 0.5 - A(1)$$

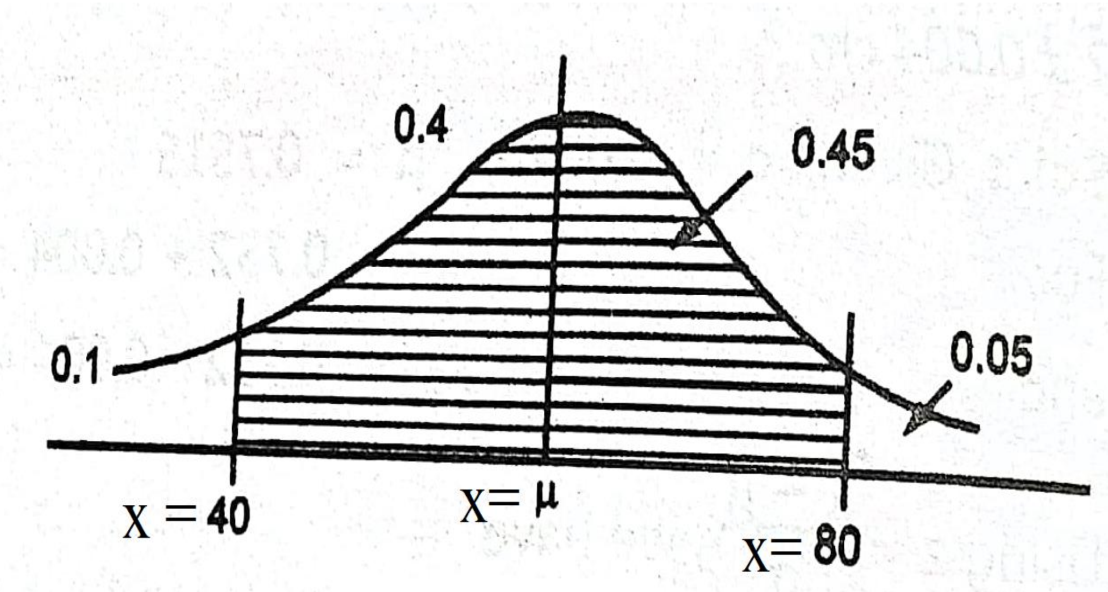
$$= 0.5 - 0.3413$$

$$= 0.1587$$

Number of tubes which will last beyond 1600 hours $= 4000 \times 0.1587 = 634.8 \sim \mathbf{635}$

Q.4 In a normal distribution 10% of items are under 40 and 5% are over 80. Find the mean and standard deviation of distribution.
Given : Area =0.4 for $z = 1.29$ & Area =0.45 for $z = 1.65$

Solution :



Given that

$$p(x<40)=\frac{10}{100}=0.1$$

$\Rightarrow x = 40$ is to the left of $x = \mu$

$$p(x>80)=\frac{5}{100}=0.05$$

$\Rightarrow x = 80$ is to the right of $x = \mu$

Probability Distribution-Normal Distribution

For $x = 40$

$$-z_1 = \frac{40-\mu}{\sigma}$$

$$\Rightarrow -1.29 = \frac{40-\mu}{\sigma}$$

$$\Rightarrow \mu - 1.29\sigma = 40....(1)$$

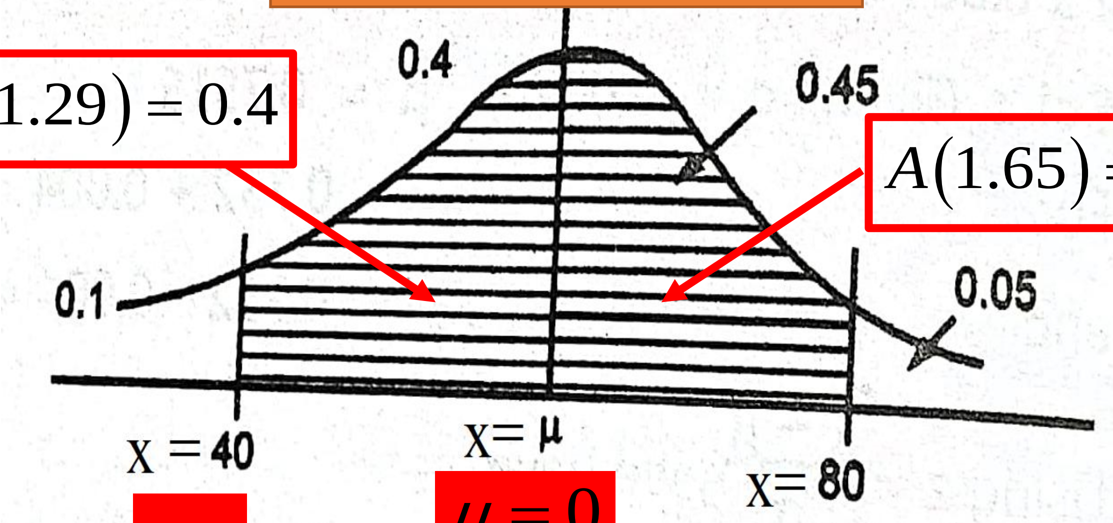
For $x = 80$

$$z_2 = \frac{80-\mu}{\sigma}$$

$$\Rightarrow 1.65 = \frac{80-\mu}{\sigma}$$

$$\Rightarrow \mu + 1.65\sigma = 80......(2)$$

After transformation



$A(1.29) = 0.4$

$A(1.65) = 0.45$

$-z_1$ $\mu = 0$ z_2

Given : Area =0.4 for $z = 1.29$ &
Area =0.45 for $z = 1.65$

$\therefore$From equation (1) & (2) $\boxed{\mu = 57.54, \sigma = 13.6}$

Q.5 In a distribution, exactly normal, 7% of the items are under 35 and 89% are under 63. Find the mean and standard deviation of the distribution.
Given : Area =0.43 for $z = 1.48$ & Area =0.39 for $z = 1.23$

Solution :

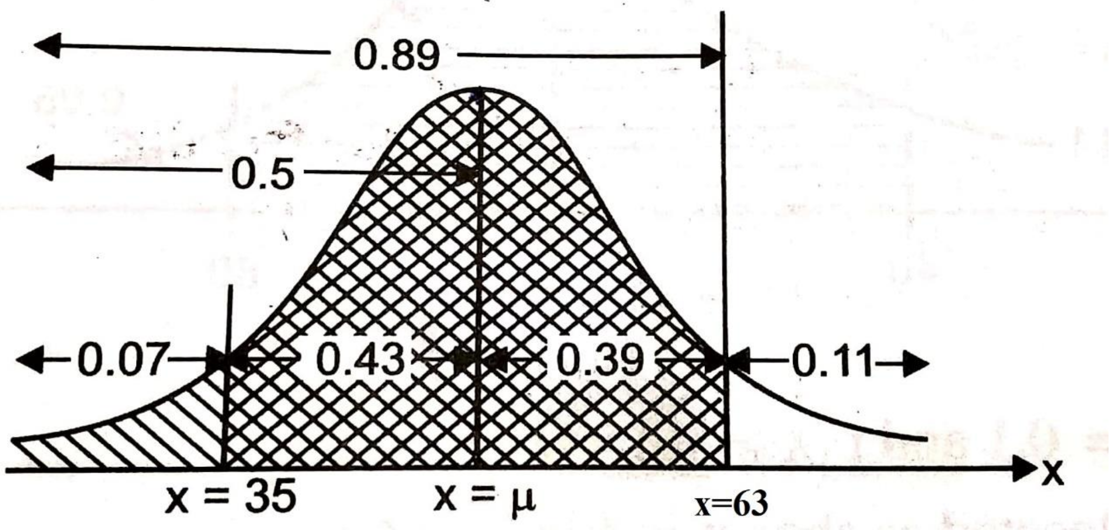
Given that

$$p(x < 35) = \frac{7}{100} = 0.07$$

$\Rightarrow x = 35$ is to the left of $x = \mu$

$$p(x < 63) = \frac{89}{100} = 0.89$$

$\Rightarrow x = 63$ is to the right of $x = \mu$



Probability Distribution-Normal Distribution

For $x = 35$

$$-z_1 = \frac{35-\mu}{\sigma}$$

$$\Rightarrow -1.48 = \frac{35-\mu}{\sigma}$$

$$\Rightarrow \mu - 1.48\sigma = 35....(1)$$

For $x = 63$

$$z_2 = \frac{63-\mu}{\sigma}$$

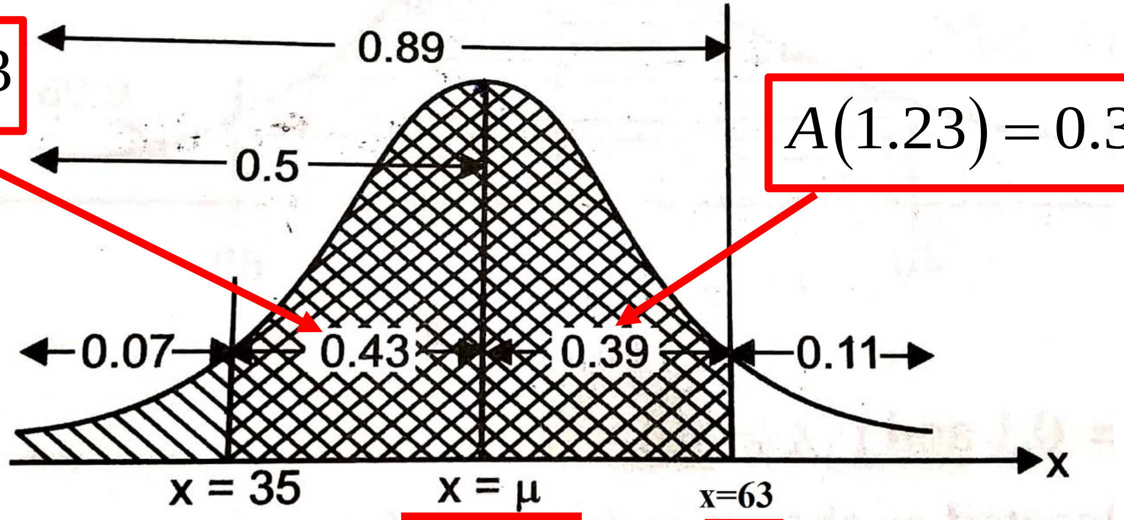
$$\Rightarrow 1.23 = \frac{63-\mu}{\sigma}$$

$$\Rightarrow \mu + 1.23\sigma = 63......(2)$$

After transformation

$A(1.48) = 0.43$

$A(1.23) = 0.39$



$-z_1$ $\mu = 0$ z_2

Given : Area =0.43 for z=1.48 &
Area =0.39 for z=1.23

$\therefore$From equation (1) & (2) $\boxed{\mu = 50.3, \sigma = 10.33}$

Pimpri Chinchwad Education Trust's

Pimpri Chinchwad College of Engineering

Department of Civil Engineering

ENGINEERING MATHEMATICS III

UNIT III STATISTICS

Mr. Dinesh Kute
(7709495376)

Statistics

Statistics

Def : Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data.

Classification of Data

1. **Ungrouped Frequency Distribution**

2. Grouped Frequency Distribution

Classification of Data

Example : Consider marks of 10 students (out of 25)

23,10,12,23,20,12,20,15,15,25

(Ungrouped Data)

Ungrouped Frequency Distribution

x	f
23	2
10	1
12	2
20	2
15	2
25	1
	$\sum_{i=1}^{6} f_i = N = 10$

Grouped Frequency Distribution

Class	f
0-5	0
5-10	0
10-15	3
15-20	2
20-25	5
	$\sum_{i=1}^{5} f_i = N = 10$

Measures of Central tendency :

Measure of central tendency is typical value around which other figures aggregate.

There are five measures of central tendency that are in common use

1) **Arithmetic average or arithmetic mean or simple mean**
2) Median
3) Mode
4) Geometric mean
5) Harmonic mean

Arithmetic mean or Average:

1. Ungrouped Data: Consider set of observations $x_1, x_2, x_3, \ldots \ldots, x_k$

$$\boxed{A.M. = \overline{x} = \frac{\sum_{i=1}^{k} x_i}{k}}$$

Arithmetic mean or Average:

2. Ungrouped Frequency distribution: Consider the frequency Distribution

x	f
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$
x_n	f_n

$$A.M = \overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

Q.1 Find mean of the following observations

(a) 480,485,492,495,500,505,515

(b)

x	32	33	34	35	36
f	4	10	13	8	5

Solution :

(a) Since, given data is ungrouped data

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i}{k} \quad \Rightarrow \bar{x} = \frac{480+485+492+495+500+505+515}{7} \quad \boxed{\Rightarrow \bar{x} = 496}$$

(b) Since, given data is in ungrouped frequency distribution

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \quad \Rightarrow \bar{x} = \frac{(32\times 4)+(33\times 10)+(34\times 13)+(35\times 8)+(36\times 5)}{4+10+13+8+5} \quad \boxed{\Rightarrow \bar{x} = 34}$$

Measures of Central Tendency

Arithmetic mean or Average:

3. Grouped Frequency distribution: Consider the frequency Distribution

Class	f	x_i
$a_1 - a_2$	f_1	$\frac{a_1 + a_2}{2}$
$a_2 - a_3$	f_2	$\frac{a_2 + a_3}{2}$
: :	: :	: :
$a_n - a_{n+1}$	f_n	$\frac{a_n + a_{n+1}}{2}$

$$\boxed{A.M. = \overline{x} = A + h\frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} = A + h\frac{\sum_{i=1}^{n} f_i u_i}{N}}$$

Where, $h : lenght\ of\ the\ interval$

$A : middle\ value\ of\ x_i$

$$u_i = \frac{x_i - A}{h}$$

Q.2 The marks obtained in paper of Mathematics are given in the following table. Find the Arithmetic mean of the distribution.

Marks Obtained	No. of students
0-10	8
10-20	20
20-30	14
30-40	16
40-50	20
50-60	25
60-70	13
70-80	10
80-90	5
90-100	2

Solution :

Measures of Central Tendency

Marks Obtained	Mid Value x	No. of students f	$u = \dfrac{x - A}{h}$	fu
0-10	5	8	-4	-32
10-20	15	20	-3	-60
20-30	25	14	-2	-28
30-40	35	16	-1	-16
40-50	45	20	0	0
50-60	55	25	1	25
60-70	65	13	2	26
70-80	75	10	3	30
80-90	85	5	4	20
90-100	95	2	5	10
Total		$\sum f = 133$		$\sum fu = -25$

$A =$ middle value of $x_i = 45$

$h = 10$

We have

$$\bar{x} = A + h\frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i}$$

$$\Rightarrow \bar{x} = 45 + 10\left(\frac{-25}{133}\right)$$

$$\boxed{\Rightarrow \bar{x} = 43.12}$$

Dispersion

Disperson

In some cases average will not give the full information. In this we check how much is scatter/ Disperse from its mean value. To do this we use following Methods of Dispersion

1. **Standard Deviation**

2. **Coefficient of Variation**

Dispersion

Standard deviation

1. Ungrouped Data: Consider set of observations $x_1, x_2, x_3, \dots \dots, x_k$

$$Standard\ Deviation = \sigma = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(x_i - \overline{x})^2}$$

2. Ungrouped Frequency Distribution : Consider the frequency distribution

x	f
x_1	f_1
x_2	f_2
:	:
:	:
x_n	f_n

$$Standard\ Deviation = \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}f_i(x_i - \overline{x})^2}$$

Dispersion

Remarks

1) $Standard\ Deviation = \sqrt{variance}$ $\quad \therefore\ \sigma^2 = variance$

2) For calculation purpose use following formulae for Standard Deviation

- Ungrouped Data

$$\boxed{\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - (\overline{x})^2}$$

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

- Ungrouped Frequency Distribution

$$\boxed{\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - (\overline{x})^2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \overline{x})^2}$$

Coefficient of Variation (C.V.) :

$$C.V. = \frac{standard\ Deviation}{A.M.} \times 100 \quad = \frac{\sigma}{\overline{x}} \times 100$$

Remarks

2) For calculation purpose use following formulae for Standard Deviation

- Ungrouped Data

$$\sigma^2 = \frac{1}{k}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 \quad = \frac{1}{k}\sum_{i=1}^{k} x_i^{\;2} - 2x_i\overline{x} + \left(\overline{x}\right)^2$$

$$= \frac{1}{k}\sum_{i=1}^{k} x_i^{\;2} - \frac{2}{k}\sum_{i=1}^{k} x_i\overline{x} + \frac{1}{k}\sum_{i=1}^{k}\left(\overline{x}\right)^2$$

$$= \frac{1}{k}\sum_{i=1}^{k} x_i^{\;2} - 2\left(\overline{x}\right)^2 + \left(\overline{x}\right)^2$$

$$= \frac{1}{k}\sum_{i=1}^{k} x_i^{\;2} + \left(\overline{x}\right)^2$$

Remarks

1. Standard Deviation gives information about distribution of data from its mean value.

2. Coefficient of Variation gives information about distribution of data from its mean value in percentage

3. If the standard deviation is 0.20 and the mean is 0.50, then the C.V. = 40%. So CV helps us see that even a lower standard deviation doesn't mean less variable data.

1. More C.V. $\Rightarrow$ More Variability$\Rightarrow$ Less Consistency

   Less C.V. $\Rightarrow$ Less Variability$\Rightarrow$ More Consistency

4. More σ $\Rightarrow$ More Variability$\Rightarrow$ Less Consistency

   Less σ $\Rightarrow$ Less Variability$\Rightarrow$ More Consistency

Examples :

Q.1 Find the standard deviation for the following data

49,63,46,59,65,52,60,54

Solution :

Standard deviation for ungrouped data is $\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - (\overline{x})^2 \qquad where,\ \overline{x} = \frac{\sum_{i=1}^{k} x_i}{k}$

x	x^2
49	2401
63	3969
46	2116
59	3481
65	4225
52	2704
60	3600
54	2916
$\sum = 448$	$\sum = 25412$

$$\therefore \sigma^2 = \frac{1}{8} \times 25412 - \left(\frac{448}{8}\right)^2 = 40.5$$

$$\therefore \sigma = 6.3640$$

Q.2 The scores obtained by two batsman A and B in 10 matches are given below. Determine who is more consistent?

Batsman A	30	44	66	62	60	34	80	46	20	38
Batsman B	34	46	70	38	55	48	60	34	45	30

Solution :

Here, we have to find coefficient of variation of Batsman A i.e. $(C.V.)_A$ & coefficient of variation of Batsman B i.e. $(C.V.)_B$

Where,

Coefficient of variation $(\mathrm{A}) = (C.V.)_A = \dfrac{\sigma_A}{(\bar{x})_A} \times 100$

Coefficient of variation $(\mathrm{B}) = (C.V.)_B = \dfrac{\sigma_B}{(\bar{x})_B} \times 100$

$$\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \left(\bar{x}\right)^2 \qquad \bar{x} = \frac{\sum_{i=1}^{k} x_i}{k}$$

Dispersion

Case I : Coefficient of variation (A) $= (C.V.)_A = \frac{\sigma_A}{(\bar{x})_A} \times 100$

Batsman A (x)	x^2
30	900
44	1936
66	4356
62	3844
60	3600
34	1156
80	6400
46	2116
20	400
38	1444
$\sum =$ 480	$\sum =$ 26152

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

$$\Rightarrow \sigma^2 = \frac{1}{10} \times 26152 - \left(\frac{480}{10}\right)^2$$

$$\Rightarrow \sigma^2 = 311.2$$

$$\Rightarrow \sigma = 17.6409$$

$$\boxed{\therefore \sigma_A = 17.6409}$$

$$\boxed{\& \ \left(\bar{x}\right)_A = \frac{480}{10} = 48}$$

$$(C.V.)_A = \frac{\sigma_A}{\left(\bar{x}\right)_A} \times 100$$

$$\therefore (C.V.)_A = \frac{17.6409}{48} \times 100$$

$$\boxed{\therefore (C.V.)_A = 36.7519}$$

Dispersion

Case II : Coefficient of variation (B) $= (C.V.)_B = \frac{\sigma_B}{(\bar{x})_B} \times 100$

Batsman B (x)	x^2
34	1156
46	2116
70	4900
38	1444
55	3025
48	2304
60	3600
34	1156
45	2025
30	900
$\sum =$ 460	$\sum =$ 22626

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

$$\Rightarrow \sigma^2 = \frac{1}{10} \times 22626 - \left(\frac{460}{10}\right)^2$$

$$\Rightarrow \sigma^2 = 146.6$$

$$\Rightarrow \sigma = 12.1078$$

$$\boxed{\therefore \sigma_B = 12.1078}$$

$$\boxed{\& \left(\bar{x}\right)_B = \frac{460}{10} = 46}$$

$$(C.V.)_B = \frac{\sigma_B}{\left(\bar{x}\right)_B} \times 100$$

$$\therefore (C.V.)_B = \frac{12.1078}{46} \times 100$$

$$\boxed{\therefore (C.V.)_B = 26.3214}$$

Dispersion

Since

$$\boxed{(C.V.)_A = 36.7519 > (C.V.)_B = 26.3214}$$

Therefore, Batsman A has more variability than Batsman B

i.e. Batsman B is more consistent than Batsman A

Q.3 Arithmetic mean and standard deviation of 30 items are 20 and 3 respectively out of these 30 items, item 22 and 15 are dropped. Find new A.M. and S.D. Calculate A.M. and S.D. if item 22 is replaced by 8 and 15 is replaced by 17.

Solution :

Let $x_1, x_2, x_3, \ldots 22, 15, \ldots . x_{30}$ be given items.

Given that,

A.M. $(\bar{x})$ and S.D. (σ) are 20 & 3 respectively

$$\therefore \bar{x} = 20 \;\&\; \sigma = 3 \ldots\ldots(1)$$

$$i.e.\ \bar{x} = \frac{\sum_{i=1}^{k} x_i}{k} \quad \& \quad \sigma^2 = \left(\frac{1}{k}\sum_{i=1}^{k} x_i^{\,2}\right) - \left(\bar{x}\right)^2$$

$$i.e.\ \frac{\sum_{i=1}^{30} x_i}{30} = 20 \quad \& \quad \frac{1}{30}\sum_{i=1}^{30} x_i^{\,2} - (20)^2 = 9$$

From equation (1)

$$i.e.\ \frac{\sum_{i=1}^{28} x_i + 22 + 15}{30} = 20 \quad \& \quad \frac{1}{30}\left[\sum_{i=1}^{28} x_i^{\,2} + 22^2 + 15^2\right] - (20)^2 = 9$$

$$i.e.\ \sum_{i=1}^{28} x_i = 563 \quad \& \quad \sum_{i=1}^{28} x_i^{\,2} = 11561$$

Case I : when item 22 & 15 are dropped

$$\therefore\ \bar{x} = \frac{\sum_{i=1}^{k} x_i}{k}$$

$$\Rightarrow \bar{x} = \frac{\sum_{i=1}^{28} x_i}{28} = \frac{563}{28} = 20.1071 \quad \boxed{\Rightarrow \bar{x} = 20.1071}$$

Find new A.M. and S.D.
Calculate A.M. and S.D. if item 22 is replaced by 8 and 15 is replaced by 17.

Dispersion

$$\sigma^2 = \left(\frac{1}{k} \sum_{i=1}^{k} x_i^{\,2} \right) - \left(\bar{x} \right)^2$$

$$i.\text{e}. \quad \sum_{i=1}^{28} x_i = 563 \quad \& \quad \sum_{i=1}^{28} x_i^{\,2} = 11561$$

$$\Rightarrow \sigma^2 = \frac{1}{28}\left[\sum_{i=1}^{28} x_i^{\,2} \right] - \left(20.1071 \right)^2$$

$$\Rightarrow \sigma^2 = \frac{1}{28}\left[11561 \right] - \left(20.1071 \right)^2$$

$$\Rightarrow \sigma^2 = 8.5974$$

$$\Rightarrow \sigma = 2.9321$$

Case II : when item 22 is replaced by 8 & 15 is replaced by 17

Now, consider the items

$$x_1, x_2, x_3, \ldots 8, 17, \ldots . x_{30}$$

Dispersion

$$\therefore\ \bar{x} = \frac{\sum_{i=1}^{k} x_i}{k} \qquad i.e.\ \sum_{i=1}^{28} x_i = 563 \ \ \& \ \sum_{i=1}^{28} x_i^{\,2} = 11561$$

$$\Rightarrow\ \bar{x} = \frac{\sum_{i=1}^{28} x_i + 8 + 17}{30} = \frac{563+8+17}{30} = 19.6 \qquad \boxed{\Rightarrow\ \bar{x} = 19.6}$$

$$and\ \ \sigma^2 = \left(\frac{1}{k}\sum_{i=1}^{k} x_i^{\,2}\right) - \left(\bar{x}\right)^2$$

$$\sigma^2 = \frac{1}{30}\left[\sum_{i=1}^{28} x_i^{\,2} + 8^2 + 17^2\right] - (19.6)^2$$

$$\sigma^2 = \frac{1}{30}\left[11561 + 8^2 + 17^2\right] - (19.6)^2$$

$$\sigma^2 = 12.9733 \qquad \boxed{\Rightarrow \sigma = 3.6019}$$

Moments, Skewness and kurtosis

Moment :

The r^{th} moment about mean $(\bar{x})$ is denoted by μ_r and defined as

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

Also, the r^{th} moment about any Number 'A' is denoted by μ'_r and defined as

$$\mu'_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^r$$

Relation between μ_r & μ'_r

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4$$

Remarks :

(1) Mean($\bar{x}$)

$$mean = \bar{x} = A + \mu_1'$$

(2) Standard Deviation(σ)

$$\text{standard deviation} = \sqrt{\mu_2}$$

(3) Coefficient of skewness(β_1)

$$\beta_1 = \frac{\mu_3^{\,2}}{\mu_2^{\,2}}$$

(4) Coefficient of kurtosis(β_2)

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}}$$

Moments, Skewness and kurtosis

Type	Description	Example	Result
Arithmetic mean	*Sum of values of a data set divided by number of values*	(1+2+2+3+4+7+9) / 7	**4**
Median	*Middle value separating the greater and lesser halves of a data set*	1, 2, 2, **3**, 4, 7, 9	**3**
Mode	*Most frequent value in a data set*	1, **2**, **2**, 3, 4, 7, 9	**2**

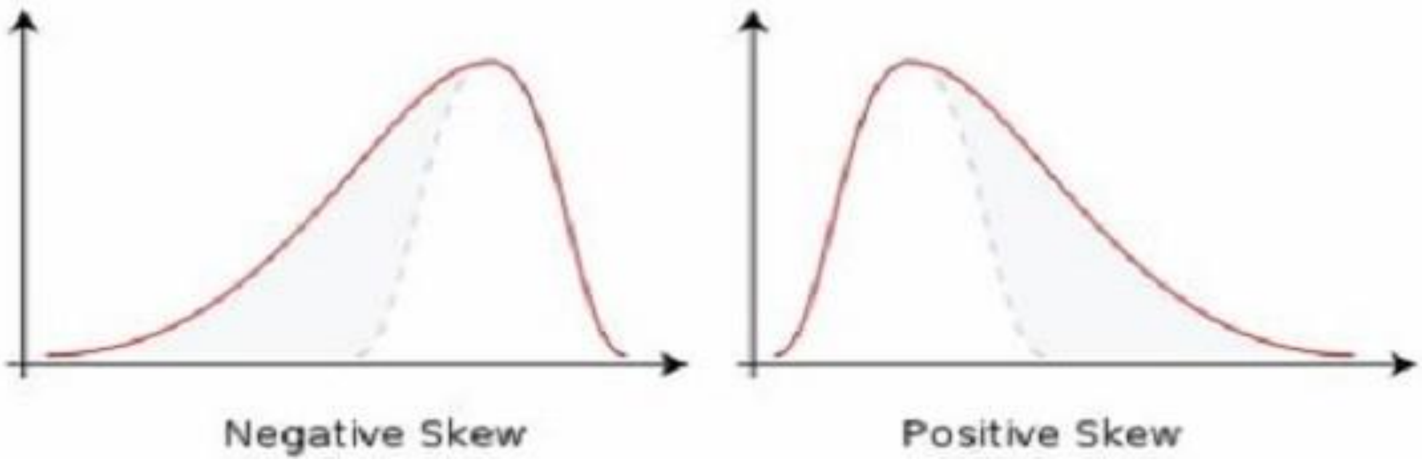
Moments, Skewness and kurtosis

Skewness : It gives imbalance and asymmetry from mean of a data distribution

There are two types of skewness

1. Positive Skewness : If the frequency curve of a distribution has a longer tail to the right of the central maximum than to the left, the distribution is said to be skewed to the right or to have positive skewed. In a positive skewed distribution, Mean > Median > Mode

2. Negative Skewness : If the frequency curve has a longer tail to the left of the central maximum than to the right, the distribution is said to be skewed to the left or to have negative skewed. In
a negatively skewed distribution, Mode > Median > Mean.



Moments, Skewness and kurtosis

Remarks :

Coefficient of skewness(β_1)

$$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^2}$$

1. $If\ \beta_1 > 0\ then\ distribution\ is\ called\ Positive\ Skew\ Distribution$

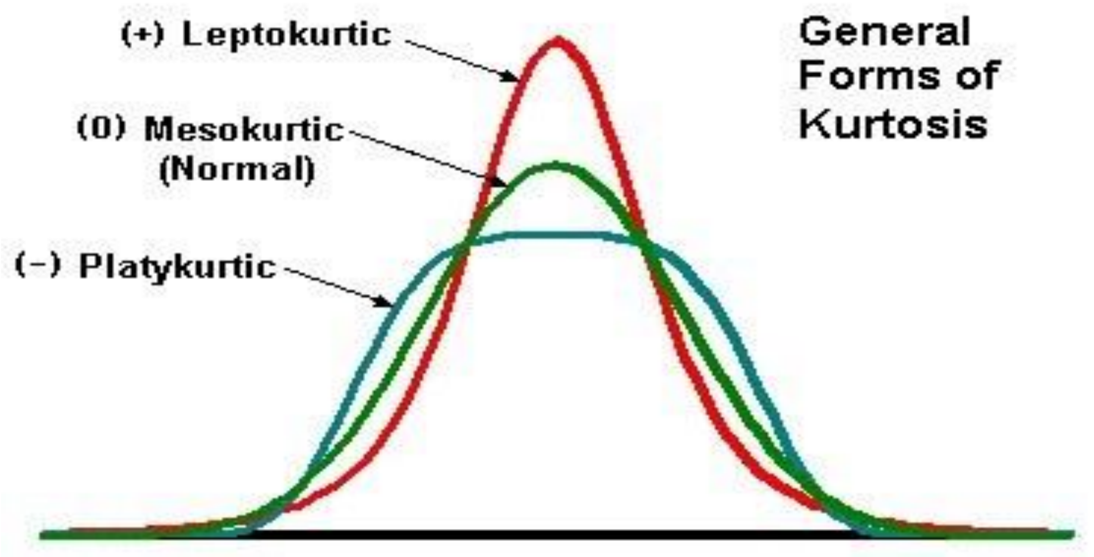
   $If\ \beta_1 < 0\ then\ distribution\ is\ called\ Negative\ Skew\ Distribution$

   $If\ \beta_1 = 0\ then\ distribution\ is\ called\ Normal\ Distribution$

Moments, Skewness and kurtosis

Kurtosis : Kurtosis is a measure of peakedness of a distribution relative to the normal distribution(or symmetrical distribution)

A distribution having a relatively high peak is called leptokurtic. A distribution which is flat topped is called platykurtic. The normal distribution which is neither very peaked nor very flat-topped is also called mesokurtic.



Remarks :

Coefficient of kurtosis(β_2)

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}}$$

2. $If\ \beta_2 > 3\ \ then\ curve\ is\ called\ Leptokurtic$

   $If\ \beta_2 = 3\ \ then\ curve\ is\ called\ Mesokurtic$

   $If\ \beta_2 < 3\ \ then\ curve\ is\ called\ Platykurtic$

Moments, Skewness and kurtosis

Q.1 The first four moments of a distribution about the value 5 are 2,20,40 and 50. From the given information obtain the first four central moments, mean, standard deviation and coefficients of skewness and kurtosis.

Solution :

Given that

$$A = 5, \mu_1' = 2, \mu_2' = 20, \mu_3' = 40, \mu_4' = 50$$

The first four central moments or moments about mean are

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 \quad = 20 - (2)^2 \quad = 16 \qquad \boxed{\Rightarrow \mu_2 = 16}$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3 = 40 - 3(2)(20) + 2(2)^3 = -64 \qquad \boxed{\Rightarrow \mu_3 = -64}$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2 \mu_2' - 3(\mu_1')^4 = 50 - 4(2)(40) + 6(2)^2(20) - 3(2)^4 = 162$$

$$\boxed{\Rightarrow \mu_4 = 162}$$

Moments, Skewness and kurtosis

Mean

$$mean = \bar{x} = A + \mu_1{}'$$

$$mean = \bar{x} = 5 + 2 \qquad \boxed{\Rightarrow mean = \bar{x} = 7}$$

Standard Deviation

$$\text{standard deviation} = \sqrt{\mu_2}$$

$$\text{standard deviation} = \sqrt{16} \qquad \boxed{\Rightarrow \text{standard deviation} = 4}$$

Coefficient of skewness

$$\beta_1 = \frac{\mu_3{}^2}{\mu_2{}^3} \qquad \Rightarrow \beta_1 = \frac{(-64)^2}{(16)^3} \qquad \boxed{\Rightarrow \beta_1 = 1}$$

Coefficient of kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2{}^2} \qquad \Rightarrow \beta_2 = \frac{162}{(16)^2} \qquad \boxed{\Rightarrow \beta_2 = 0.6328}$$

Q.2 Find the four moments about the mean of the following

x	61	64	67	70	73
f	5	18	42	27	8

also calculate β_1 & β_2.

Solution :

The moment about mean $\bar{x}$ is given below $\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r$

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)$$

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^3$$

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2$$

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^4$$

Moments, Skewness and kurtosis

x	f	fx	$f(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
61	5	305	-32.25	208.0125	-1341.6806	8653.84
64	18	1152	-62.1	214.245	-739.1453	2550.051
67	42	2814	-18.9	8.505	-3.8273	1.722263
70	27	1890	68.85	175.5675	447.6971	1141.628
73	8	584	44.4	246.42	1367.6310	7590.352
$\sum = 335$	$N = \sum = 100$	$\sum = 6745$	$\sum = 0$	$\sum = 852.75$	$\sum = -269.3250$	$\sum = 19937.5931$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=}^{n} f_i} = \frac{6745}{100} = 67.45$$

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right) = \frac{1}{100}\times 0 = 0$$

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2 = \frac{1}{100}\times 852.75 = 8.5275$$

Moments, Skewness and kurtosis

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^3 \quad = \frac{1}{100}\times\left(-269.3250\right) = -2.6933$$

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^4 = \frac{1}{100}\times\left(19937.5931\right) = 199.3759$$

Coefficient of skewness

$$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^2} \Rightarrow \beta_1 = \frac{\left(-2.6933\right)^2}{\left(8.5275\right)^2} \quad \boxed{\Rightarrow \beta_1 = 0.0998}$$

Coefficient of kurtosis

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2} \Rightarrow \beta_2 = \frac{199.3759}{\left(8.5275\right)^2} \quad \boxed{\Rightarrow \beta_2 = 2.7418}$$

Correlation

Correlation :

To measure linear relationship between two variables, Karl Pearson developed a formula called correlation coefficient.

Coefficient of Correlation between two variables x and y denoted by $r(x,y)$ and defined as

$$\boxed{r(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y}}$$

Where

1. Ungrouped Data

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

$$\boxed{cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}.\overline{y}}$$

$$\boxed{\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2} \quad \boxed{\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - (\overline{y})^2}$$

2. Ungrouped Frequency Distribution

$$cov(x,y) = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \overline{x})(y_i - \overline{y})$$

$$\boxed{cov(x,y) = \frac{1}{N}\sum_{i=1}^{n} f_i x_i y_i - \overline{x}.\overline{y}}$$

$$\boxed{\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - (\overline{x})^2} \quad \boxed{\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{n} f_i y_i^2 - (\overline{y})^2}$$

Correlation

Remarks :

1. $-1 < r < 1$

2. If r is close to 0, it means there is no relationship between the variables. If r is positive, it means that as one variable gets larger the other gets larger. If r is negative it means that as one gets larger, the other gets smaller

EXAMPLES

Q.1 Find the coefficient of correlation for the following table

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Solution :

We know, Karl Pearson's coefficient of correlation $r(x, y)$

$$r(x, y) = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y}$$

Correlation

$$where,\ \operatorname{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\left(\bar{x}.\bar{y}\right)$$

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

$$\sigma_y{}^2=\frac{1}{n}\sum_{i=1}^{n}y_i{}^2-\left(\bar{y}\right)^2$$

x	y	xy	x^2	y^2
10	18	180	100	324
14	12	168	196	144
18	24	432	324	576
22	6	132	484	36
26	30	780	676	900
30	36	1080	900	1296
$\sum=120$	$\sum=126$	$\sum=2772$	$\sum=2680$	$\sum=3276$

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{120}{6}=20 \qquad \boxed{\Rightarrow \bar{x}=20}$$

$$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{126}{6}=21 \qquad \boxed{\Rightarrow \bar{y}=21}$$

Correlation

$$\operatorname{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\left(\bar{x}.\bar{y}\right)$$

$$\operatorname{cov}(x,y)=\frac{1}{6}\times 2772-\left(20\times 21\right)$$

$$\boxed{\operatorname{cov}(x,y)=42}$$

x	y	xy	x^2	y^2
10	18	180	100	324
14	12	168	196	144
18	24	432	324	576
22	6	132	484	36
26	30	780	676	900
30	36	1080	900	1296
$\Sigma = 120$	$\Sigma = 126$	$\sum=2772$	$\sum=2680$	$\sum=3276$

$$\boxed{\Rightarrow \bar{x}=20} \qquad \boxed{\Rightarrow \bar{y}=21}$$

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

$$\sigma_x{}^2=\frac{1}{6}\times 2680-\left(20\right)^2$$

$$\sigma_x{}^2=46.6667 \qquad \boxed{\Rightarrow \sigma_x=6.8313}$$

$$\sigma_y{}^2=\frac{1}{n}\sum_{i=1}^{n}y_i{}^2-\left(\bar{y}\right)^2 \qquad \Rightarrow \sigma_y{}^2=\frac{1}{6}\times 3276-\left(21\right)^2 \qquad \Rightarrow \sigma_y{}^2=105 \qquad \boxed{\Rightarrow \sigma_y=10.2470}$$

Correlation

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y}$$

$$\boxed{\operatorname{cov}(x,y)=42}$$

$$\boxed{\Rightarrow \sigma_x = 6.8313}$$

$$\boxed{\Rightarrow \sigma_y = 10.2470}$$

$$r(x,y)=\frac{42}{(6.8313)(10.2470)}$$

$$\boxed{r(x,y)=0.6}$$

EXAMPLES

Q.2 Find the coefficient of correlation for the following table

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

Solution :

We know, Karl Pearson's coefficient of correlation $r(x,y)$

$$\boxed{\Rightarrow r(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y}}$$

Correlation

$$where,\ \operatorname{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\left(\bar{x}.\bar{y}\right)$$

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

$$\sigma_y{}^2=\frac{1}{n}\sum_{i=1}^{n}y_i{}^2-\left(\bar{y}\right)^2$$

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum = 30$	$\sum = 40$	$\sum = 214$	$\sum = 220$	$\sum = 340$

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{30}{5}=6$$

$$\boxed{\Rightarrow \bar{x}=6}$$

$$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{40}{5}=8$$

$$\boxed{\Rightarrow \bar{y}=8}$$

Correlation

$$\text{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-(\bar{x}.\bar{y})$$

$$\text{cov}(x,y)=\frac{1}{5}\times 214-(6\times 8)$$

$$\boxed{\text{cov}(x,y)=-5.2}$$

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\Sigma = 30$	$\Sigma = 40$	$\sum = 214$	$\sum = 220$	$\sum = 340$

$$\boxed{\Rightarrow \bar{x}=6} \qquad \boxed{\Rightarrow \bar{y}=8}$$

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

$$\sigma_x{}^2=\frac{1}{5}\times 220-(6)^2$$

$$\sigma_x{}^2=8 \qquad \boxed{\Rightarrow \sigma_x=2.8284}$$

$$\sigma_y{}^2=\frac{1}{n}\sum_{i=1}^{n}y_i{}^2-(\bar{y})^2 \quad \Rightarrow \sigma_y{}^2=\frac{1}{5}\times 340-(8)^2 \quad \Rightarrow \sigma_y{}^2=4 \qquad \boxed{\Rightarrow \sigma_y=2}$$

Correlation

$$r(x,y)=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y}$$

$$\boxed{\operatorname{cov}(x,y)=-5.2}$$

$$\boxed{\Rightarrow \sigma_x=2.8284}$$

$$\boxed{\Rightarrow \sigma_y=2}$$

$$r(x,y)=\frac{-5.2}{(2.8284)(2)}$$

$$\boxed{r(x,y)=-0.9192}$$

Q.3 Find the coefficient of correlation for the following table

$$n=20,\sum x=40,\sum x_i^{\,2}=190,\sum y_i=40,\sum y_i^{\,2}=200,\sum x_iy_i=150,$$

Solution :

We know that

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{40}{20}=2 \quad \& \quad \bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{40}{20}=2$$

Correlation

Q.3 Find the coefficient of correlation for the following table

$n = 20, \sum x = 40, \sum x_i^2 = 190, \sum y_i = 40, \sum y_i^2 = 200, \sum x_i y_i = 150,$

$\boxed{\bar{x} = 2}$ $\boxed{\bar{y} = 2}$

$$r(x, y) = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y}$$

$$where,\ \operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{x}.\bar{y})$$

$$= \frac{1}{20} \times 150 - (2 \times 2)$$

$$= 3.5$$

$$\boxed{\Rightarrow \operatorname{cov}(x, y) = 3.5}$$

$$\sigma_x^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\ 2} - (\bar{x})^2 = \frac{1}{20} \times 190 - (2)^2 = 5.5 \quad \boxed{\Rightarrow \sigma_x = 2.3452}$$

$$\sigma_y^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{\ 2} - (\bar{y})^2 = \frac{1}{20} \times 200 - (2)^2 = 6 \quad \boxed{\Rightarrow \sigma_y = 2.4495}$$

$$\Rightarrow r(x, y) = \frac{3.5}{(2.3452 \times 2.4495)}$$

$$\boxed{\Rightarrow r(x, y) = 0.6093}$$

Regression :

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables.

The line of regression gives best estimate for the value of one variable for some specified value of other variable

Regression line y on x

$$y - \bar{y} = b_{yx}(x - \bar{x}) \qquad \text{where } b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{\operatorname{cov}(x,y)}{\sigma_x . \sigma_y}\frac{\sigma_y}{\sigma_x} \qquad \boxed{\Rightarrow b_{yx} = \frac{\operatorname{cov}(x,y)}{\sigma_x{}^2}}$$

Regression line x on y

$$x - \bar{x} = b_{xy}(y - \bar{y}) \qquad \text{where } b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x,y)}{\sigma_x . \sigma_y}\frac{\sigma_x}{\sigma_y} \qquad \boxed{\Rightarrow b_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_y{}^2}}$$

where $\boxed{\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\bar{x}.\bar{y}\right)}$ $\boxed{\sigma_x{}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i{}^2 - \left(\bar{x}\right)^2}$ $\boxed{\sigma_y{}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i{}^2 - \left(\bar{y}\right)^2}$

Regression

Remarks :

1) $b_{yx} = r\frac{\sigma_y}{\sigma_x}$, $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ are called regression coefficients

2) $r = \sqrt{b_{yx}.b_{xy}}$

3) If x is given and to find corresponding value of y ,use regression line of y on x

4) If y is given and to find corresponding value of x ,use regression line of x on y

5) If $b_{xy} > 0$ and $b_{yx} > 0$ then $r > 0$

6) If $b_{xy} < 0$ and $b_{yx} < 0$ then $r < 0$

7) Point $(\bar{x}, \bar{y})$ satisfies the equations of lines of regression

$$\boxed{y - \bar{y} = b_{yx}\left(x - \bar{x}\right)} \qquad \boxed{x - \bar{x} = b_{xy}\left(y - \bar{y}\right)}$$

Regression

Q.1 Obtain regression lines for the following data :

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

Solution :

Case I : Regression y on x

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

where

$$b_{yx}=\frac{\text{cov}\left(x,y\right)}{\sigma_x{}^2}$$

Case II : Regression x on y

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

where

$$b_{xy}=\frac{\text{cov}\left(x,y\right)}{\sigma_y{}^2}$$

and

$$\boxed{\text{cov}\left(x,y\right)=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\left(\bar{x}.\bar{y}\right)}$$

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

$$\boxed{\sigma_y{}^2=\frac{1}{n}\sum_{i=1}^{n}y_i{}^2-\left(\bar{y}\right)^2}$$

Regression

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum = 30$	$\sum = 40$	$\sum = 214$	$\sum = 220$	$\sum = 340$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{30}{5} = 6$$

$$\boxed{\Rightarrow \bar{x} = 6}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{40}{5} = 8$$

$$\boxed{\Rightarrow \bar{y} = 8}$$

$$\boxed{\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{x}.\bar{y})}$$

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

$$\boxed{\sigma_y{}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i{}^2 - (\bar{y})^2}$$

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{x}.\bar{y}) = \frac{1}{5} \times 214 - (6 \times 8) = -5.2$$

Regression

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum = 30$	$\sum = 40$	$\sum = 214$	$\sum = 220$	$\sum = 340$

$$\boxed{\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{x}.\bar{y})}$$

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

$$\boxed{\sigma_y^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{\ 2} - (\bar{y})^2}$$

$$\boxed{\Rightarrow \bar{x} = 6} \quad \boxed{\Rightarrow \bar{y} = 8}$$

$$\boxed{\Rightarrow \operatorname{cov}(x,y) = -5.2}$$

$$\sigma_x^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\ 2} - (\bar{x})^2 = \frac{1}{5}\times 220 - (6)^2 = 8$$

$$\boxed{\Rightarrow \sigma_x^{\ 2} = 8}$$

$$\sigma_y^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{\ 2} - (\bar{y})^2 = \frac{1}{5}\times 340 - (8)^2 = 4$$

$$\boxed{\Rightarrow \sigma_y^{\ 2} = 4}$$

$$b_{yx} = \frac{\operatorname{cov}(x,y)}{\sigma_x^{\ 2}} = \frac{-5.2}{8} = -0.65 \quad \boxed{\Rightarrow b_{yx} = -0.65}$$

$$b_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_y^{\ 2}} = \frac{-5.2}{4} = -1.3 \quad \boxed{\Rightarrow b_{xy} = -1.3}$$

Regression

Case I : Regression y on x

$$\Rightarrow b_{yx} = -0.65$$

$$\Rightarrow b_{xy} = -1.3$$

$$\Rightarrow \bar{x} = 6 \qquad \Rightarrow \bar{y} = 8$$

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

$$y - 8 = -0.65(x - 6)$$

$$y - 8 = -0.65x + 3.9$$

$$\boxed{y = -0.65x + 11.9}$$

Case II : Regression x on y

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

$$x - 6 = -1.3(y - 8)$$

$$x - 6 = -1.3y + 10.4$$

$$\boxed{x = -1.3y + 16.4}$$

Q.2 Obtain regression lines for the following data :

x	10	14	19	26	30	34	39
y	12	16	18	26	29	35	38

and estimate y for $x = 14.5$ and x for $y = 29.5$

Solution :

Case I : Regression y on x

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

where

$$b_{yx} = \frac{\operatorname{cov}(x, y)}{\sigma_x{}^2}$$

Case II : Regression x on y

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

where

$$b_{xy} = \frac{\operatorname{cov}(x, y)}{\sigma_y{}^2}$$

and

$$\boxed{\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\bar{x}.\bar{y}\right)}$$

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

$$\boxed{\sigma_y{}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i{}^2 - \left(\bar{y}\right)^2}$$

Regression

x	y	xy	x^2	y^2
10	12	120	100	144
14	16	224	196	256
19	18	342	361	324
26	26	676	676	676
30	29	870	900	841
34	35	1190	1156	1225
39	38	1482	1521	1444
$\sum = 172$	$\sum = 174$	$\sum = 4904$	$\sum = 4910$	$\sum = 4910$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{172}{7} = 24.5714 \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{174}{7} = 24.8571$$

$$\boxed{\Rightarrow \bar{x} = 24.5714} \qquad \boxed{\Rightarrow \bar{y} = 24.8571}$$

$$\boxed{\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\bar{x}.\bar{y}\right)}$$

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

$$\boxed{\sigma_y^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{\ 2} - \left(\bar{y}\right)^2}$$

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\bar{x}.\bar{y}\right) = \frac{1}{7} \times 4904 - \left(24.5714 \times 24.8571\right)$$

Regression

$$\boxed{\Rightarrow \operatorname{cov}(x,y) = 89.7977}$$

x	y	xy	x^2	y^2
10	12	120	100	144
14	16	224	196	256
19	18	342	361	324
26	26	676	676	676
30	29	870	900	841
34	35	1190	1156	1225
39	38	1482	1521	1444
$\sum = 172$	$\sum = 174$	$\sum = 4904$	$\sum = 4910$	$\sum = 4910$

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - (\bar{x}.\bar{y})$$

$$\sigma_x^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\ 2} - (\bar{x})^2 = \frac{1}{7}\times 4910 - (24.5714)^2$$

$$\boxed{\Rightarrow \sigma_x^{\ 2} = 97.6749}$$

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

$$\sigma_y^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{\ 2} - (\bar{y})^2 = \frac{1}{7}\times 4910 - (24.8571)^2$$

$$\boxed{\Rightarrow \sigma_y^{\ 2} = 83.5532}$$

$$\sigma_y^{\ 2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{\ 2} - (\bar{y})^2$$

$$b_{yx} = \frac{\operatorname{cov}(x,y)}{\sigma_x^{\ 2}} = \frac{89.7977}{97.6749} \qquad \boxed{\Rightarrow b_{yx} = 0.9194}$$

$$\boxed{\Rightarrow \bar{x} = 24.5714}$$

$$b_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_y^{\ 2}} = \frac{89.7977}{83.5532} \qquad \boxed{\Rightarrow b_{xy} = 1.0747}$$

$$\boxed{\Rightarrow \bar{y} = 24.8571}$$

Regression

$$\Rightarrow b_{yx} = 0.9194$$

$$\Rightarrow b_{xy} = 1.0747$$

$$\Rightarrow \bar{x} = 24.5714$$

$$\Rightarrow \bar{y} = 24.8571$$

Case I : Regression y on x

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

$$y - 24.8571 = 0.9194\left(x - 24.5714\right)$$

$$y - 24.8571 = 0.9194x - 22.5909$$

$$\boxed{y = 0.9194x + 2.2662}$$

and estimate

y for $x = 14.5$ and x for $y = 29.5$

For $x = 14.5$

$$y = 0.9194\left(14.5\right) + 2.2662$$

$$\boxed{y = 15.5975}$$

Regression

Case II : Regression x on y

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

$$x-24.5714=1.0747\left(y-24.8571\right)$$

$$x-24.5714=1.0747y-26.7139$$

$$\boxed{x=1.0747y-2.1425}$$

For y = 29.5

$$x=1.0747\left(29.5\right)-2.1425$$

$$\boxed{x=29.5612}$$

$$\boxed{\Rightarrow b_{yx}=0.9194}$$

$$\boxed{\Rightarrow b_{xy}=1.0747}$$

$$\boxed{\Rightarrow \bar{x}=24.5714}$$

$$\boxed{\Rightarrow \bar{y}=24.8571}$$

and estimate

y for $x=14.5$ and x for $y=29.5$

Regression

Q.3 The regression equations are $8x-10y+66=0$ and $40x-18y=214$.The value of variance of x is 9. Find

a) The mean values of x and y

b) The correlation x and y and

c) The standard deviation of y

Solution

The regression lines are $8x-10y+66=0$ and $40x-18y=214$

Case I : Mean values of x and y

Since, we know that the point $(\bar{x},\bar{y})$ satisfies both regression lines

$$8\bar{x}-10\bar{y}+66=0 \text{ and } 40\bar{x}-18\bar{y}=214$$

$$\boxed{\bar{x}=13 \text{ and } \bar{y}=17}$$

Regression

Case II : Coefficient of correlation

$$8x - 10y + 66 = 0$$

$$40x - 18y = 214$$

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

$$\boxed{-1 < r < 1}$$

Regression y on x : $y - \bar{y} = b_{yx}\left(x - \bar{x}\right)$

Regression x on y : $x - \bar{x} = b_{xy}\left(y - \bar{y}\right)$

To find regression of y on x and regression of x on y

$8x - 10y + 66 = 0$	$r = \sqrt{b_{xy} . b_{yx}}$	$40x - 18y = 214$
$y = \frac{8}{10}x + \frac{66}{10} \quad \Rightarrow b_{yx} = \frac{8}{10} = 0.8$	1) $r = \sqrt{0.8 \times 0.45}$ $\boxed{r = 0.6}$	$\cancel{y = \frac{40}{18}x - \frac{214}{18} \Rightarrow b_{yx} = \frac{40}{18} = 2.2222}$
$\cancel{x = \frac{10}{8}y - \frac{66}{8} \Rightarrow b_{xy} = \frac{10}{8} = 1.25}$	$\cancel{2)\ r = \sqrt{1.25 \times 2.2222}}$ $\cancel{r = 1.6667}$	$x = \frac{18}{40}y + \frac{214}{40} \Rightarrow b_{xy} = \frac{18}{40} = 0.45$

Regression

$8x-10y+66=0$	$r=\sqrt{b_{xy}.b_{yx}}$	$40x-18y=214$
$y=\frac{8}{10}x+\frac{66}{10} \Rightarrow b_{yx}=\frac{8}{10}=0.8$	1) $r=\sqrt{0.8\times0.45}$ $\boxed{r=0.6}$	
		$x=\frac{18}{40}y+\frac{214}{40} \Rightarrow b_{xy}=\frac{18}{40}=0.45$

Regression y on x :

$$y=\frac{8}{10}x+\frac{66}{10}$$

$$\Rightarrow b_{yx}=\frac{8}{10}=0.8$$

Regression x on y :

$$x=\frac{18}{40}y+\frac{214}{40}$$

$$\Rightarrow b_{xy}=\frac{18}{40}=0.45$$

$$r=\sqrt{0.8\times0.45} \qquad \boxed{\Rightarrow r=0.6}$$

Case II : standard deviation of y : (σ_x)

standard deviation $=\sqrt{\text{variance}}$ $\quad \Rightarrow \sigma=\sqrt{\text{var}}$

Given that $var(x)=9$

$$\Rightarrow \sigma_x=\sqrt{\text{var}(x)}$$

$$\Rightarrow \sigma_x=\sqrt{9} \qquad \boxed{\Rightarrow \sigma_x=3}$$

Regression

Q.4 If the two lines of regression are $9x+y-\lambda=0$ and $4x+y=\mu$ and the means of x & y are 2 & -3 respectively, find the values of λ,μ and the coefficient of correlation between x & y.

Solution

The regression lines are $9x+y=\lambda$ and $4x+y=\mu$

Since, we know that the point $(\bar{x},\bar{y})$ satisfies both regression lines

$$\therefore 9\bar{x}+\bar{y}=\lambda \text{ and } 4\bar{x}+\bar{y}=\mu$$

Given that : $\bar{x}=2$ and $\bar{y}=-3$

$$\therefore 9(2)+(-3)=\lambda \text{ and } 4(2)+(-3)=\mu$$

$$\boxed{\lambda=15 \text{ and } \mu=5}$$

$\therefore$The regression lines are $9x+y=15$ and $4x+y=5$

Regression

Case II : Coefficient of correlation

$\boxed{\lambda = 15 \text{ and } \mu = 5}$

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

Regression y on x : $y - \bar{y} = b_{yx}\left(x - \bar{x}\right)$

$\boxed{-1 < r < 1}$

Regression x on y : $x - \bar{x} = b_{xy}\left(y - \bar{y}\right)$

To find regression of y on x and regression of x on y

$9x + y = 15$	$r = \sqrt{b_{xy}.b_{yx}}$	$4x + y = 5$
$\cancel{y = -9x + 15 \Rightarrow b_{yx} = -9}$	1) $\cancel{r = \sqrt{(-9)\times(-0.25)}}$ $\cancel{r = 1.5}$	$y = -4x + 5 \Rightarrow b_{yx} = -4$
$x = \frac{-1}{9}y + \frac{15}{9} \Rightarrow b_{xy} = \frac{-1}{9} = -0.1111$	2) $r = \sqrt{(-0.1111)\times(-4)}$ $\boxed{r = 0.6666}$	$\cancel{x = \frac{-1}{4}y + \frac{5}{4} \Rightarrow b_{xy} = \frac{-1}{4} = -0.25}$

Regression

$9x+y=15$	$r=\sqrt{b_{xy}.b_{yx}}$	$4x+y=5$
		$y=-4x+5 \Rightarrow b_{yx}=-4$
$x=\frac{-1}{9}y+\frac{15}{9} \Rightarrow b_{xy}=\frac{-1}{9}=-0.1111$	2) $r=\sqrt{(-0.1111)\times(-4)}$ $\boxed{r=0.6666}$	

Regression y on x :

$$x = \frac{-1}{9}y + \frac{15}{9}$$

$$\Rightarrow b_{xy} = \frac{-1}{9} = -0.1111$$

Regression x on y :

$$y = -4x + 5$$

$$\Rightarrow b_{yx} = -4$$

$$r = \sqrt{(-0.1111)\times(-4)}$$

$$r = 0.6666$$

Since both regression coefficients b_{xy} and b_{yx} are negative

$\therefore$ we take

$$\boxed{r = -0.6666}$$