STATISTICAL TECHNIQUE-II

Probability

Probability : Probability measures the degree of certainity or uncertainity of the occurance of events.

If event A happens in m ways and fails in n ways then probability of happening A, $p = \frac{m}{m+n}$.

and probability of failure $q = \frac{n}{m+n}$.

Obviously, $p + q = \frac{m}{m+n} + \frac{n}{m+n} = 1.$

Law of Addition : If $p_1, p_2, \ldots, p_n$ be probability of mutually exclusive events then probability of happening any of these events,

$$P = p_1 + p_2 + \ldots + p_n.$$

Proof. Let $A_1, A_2, \ldots, A_n$ be events with probabilities $p_1, p_2, \ldots, p_n$.

$$P(A_1 + A_2 + \ldots + A_n) = \frac{m_1 + m_2 + \ldots + m_n}{n}$$

$$= \frac{m_1}{n} + \frac{m_2}{n} + \ldots + \frac{m_n}{n}$$

$$= p_1 + p_2 + \ldots + p_n.$$

Law of Multiplication : If there are two independent events with known probabilities then probability that both will happen is equal to the product of their probabilities.

$$P(A B) = P(A) \cdot P(B).$$

Example 1. A bag contains 6 black and 5 red balls. Find probability of drawing *two balls of the same colour.*

Solution. *Probability of drawing 2 black balls*

$$= \frac{{}^6C_2}{{}^{11}C_2} = \frac{15}{55} = \frac{3}{11}$$

Probability of drawing 2 real balls

$$= \frac{{}^5C_2}{{}^{11}C_2} = \frac{10}{55} = \frac{2}{11}$$

$\therefore$ Probability of drawing 2 ball of same colour

$$= \frac{3}{11} + \frac{2}{11} = \frac{5}{11} \qquad \textit{Ans. } \frac{5}{11}$$

Example 2. An urn contains 2 red, 3 blue and 4 white balls. 3 balls are drawn at random. Find the probability that all the three balls are of different colour.

Solution. The three ball may be of different colour implies one ball is red, one ball is blue and one ball is white.

Probability of 3 balls be of different colour

$$= \frac{{}^3C_1 \times {}^3C_1 \times {}^4C_1}{{}^{10}C_3} = \frac{3 \times 3 \times 4}{120} = \frac{3}{10} \qquad \textit{Ans. } \frac{3}{10}$$

Binomial Distribution

Binomial Probability Distribution was discovered by Swiss Mathematician J.J. Bernoulli in year 1700.

The occurance of an event is called 'a success' and non-occurance of an event is called 'a failure'. Let x be number of successes in n. trials. If p be probability of success and q that of failure in a single trial then $p + q = 1$.

Probability of r successes out of n trials,

$$P(r) = {}^nC_r \times \text{Prob. of } (r \text{ successes and } n - r \text{ failure})$$

$$\Rightarrow P(r) = {}^nC_r \cdot p^r q^{n-r}$$

It is called Binomial Probability Distribution.

Recurrence formula for Binomial Distribution :

By Bonomial Distribution, we have

$$P(r) = {}^nC_r p^r q^{n-r} = \frac{n!}{r!(n-r)!} p^r \cdot q^{n-r} \qquad \ldots(1)$$

$$\text{Also, } P(r+1) = {}^nC_{r+1} p^{r+1} \cdot q^{n-r-1} = \frac{n!}{(r+1)!(n-r-1)!} p^{r+1} \cdot q^{n-r-1} \qquad \ldots(2)$$

$$\therefore \frac{P(r+1)}{P(r)} = \frac{n!}{(r+1)!(n-r-1)!} \times \frac{r!(n-r)!}{n!} \cdot \frac{p^{r+1} \cdot q^{n-r-1}}{p^r \cdot q^{n-r}}$$

$$\Rightarrow \frac{P_{(r+1)}}{P(r)} = \left(\frac{n-r}{r+1}\right) \cdot \frac{p}{q}$$

$$\Rightarrow P(r+1) = \left(\frac{n-r}{r+1}\right) \frac{p}{q} P(r)$$

Mean of the Binomial Distribution

By Binomial Distribution.

$$P(r) = {}^nc_r\, p^r \cdot q^{n-r}$$

and Mean, $\mu = \sum_{r=0}^{n} rP(r) = \sum_{r=0}^{n} r \cdot {}^nc_r p^r q^{n-r}$

$$\Rightarrow \mu = 0 + 1 \cdot {}^nc_1 pq^{n-1} + 2 \cdot {}^nc_2 p^2 q^{n-2} + 3 \cdot {}^nc_3 p^3 q^{n-3} + \ldots + r \cdot {}^nc_r p^r q^{n-r} + \ldots + np^n$$

$$= npq^{n-1} + \frac{n(n-1)}{1!} p^2 q^{n-2} + \frac{n(n-1)(n-2)}{2!} p^3 q^{n-3} + \ldots + n \cdot p^n$$

$$= np\left[q^{n-1} + (n-1)p \cdot q^{n-2} + \frac{(n-1)(n-1)}{2!} p^2 q^{n-3} + \ldots + p^{n-1}\right]$$

$$= np\,[q+p]^{n-1}$$

$$= np \cdot (1)^{n-1} = np$$

$\therefore$ Mean, $\mu = np$

Variance of Binomial Distribution

Binomial distribution

Mean $\mu = np$

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

Variance, $\sigma^2 = \sum_{r=0}^{n} r^2 \cdot P(r) - \mu^2$

$$\Rightarrow \sigma^2 = \sum_{r=0}^{n} r^2 \cdot {}^nc_r p^r q^{n-r} - (np)^2$$

$$\Rightarrow \sigma^2 = \sum_{r=0}^{n} [r + r(r-1)]\, {}^nc_r p^r q^{n-r} - n^2p^2$$

$$\Rightarrow \sigma^2 = \sum_{0}^{n} r \cdot {}^nc_r \cdot p^r q^{n-r} + \sum_{0}^{n} r(r-1)\, {}^nc_r p^r q^{n-r} - n^2p^2$$

$$\Rightarrow \sigma^2 = \mu + 2! \cdot {}^nc_2\, p^2 q^{n-2} + 3!\, {}^nc_3 p^r q^{n-r} + \ldots + n(n-1) \cdot p^n - n^2p^2$$

$$\Rightarrow \sigma^2 = np + n(n-1)p^2 q^{n-2} + n(n-1)(n-2)p^3 q^{n-3} + \ldots + n(n-1)p^n - n^2p^2$$

$$\Rightarrow \sigma^2 = np + n(n-1)^2 p\,[q^{n-2} + (n-2)p \cdot q^{n-3} + \ldots + p^{n-2}] - n^2p^2$$

$$\Rightarrow \sigma^2 = np + n(n-1)\, p^2\, [q+p]^{n-2} - n^2p^2$$

$$\Rightarrow \sigma^2 = np + n(n-1)\, p^2 \cdot (1)^{n-2} - n^2p^2$$

$$\Rightarrow \sigma^2 = np + n^2p^2 - np^2 - n^2p^2$$

$$\Rightarrow \sigma^2 = np - np^2$$

$$\Rightarrow \sigma^2 = np\,(1-p).$$

$\Rightarrow$ Variance, $\sigma^2 = npq$

S. D., $\sigma = \sqrt{npq}$.

Example 1. Five dice are thrown 2187 times. How many times at least 3 dice will show a four or two ?

Solution. Probability of getting a four or two with one die, $p = \frac{2}{6} = \frac{1}{3}$

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

Given $n = 5$, $N = 2187$

$\therefore$ Number of times at least 3 dice showing a four or two

$= N \cdot P(r \geq 3)$

$= 2187 \cdot [P(3) + P(4) + P(5)]$

$= 2187\, [{}^nc_3 p^3 q^{n-3} + {}^nc_4 p^4 q^{n-4} + {}^nc_5 p^5 q^{n-5}]$

$$= 2187\left[{}^5c_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2 + {}^5c_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^1 + {}^5c_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^0\right]$$

$$= 2187\left[5 \cdot \frac{(2)^2}{3^5} + 5 \cdot \frac{(2)^3}{3^5} + \frac{(2)^4}{3^5}\right]$$

$$= \frac{(3)^7}{3^5} \cdot [20 + 40 + 16]$$

$= 9\,[76] = 684$ **Answer.**

Example 2. If 10% of bolts produced by a machine are defective. Find the probability that out of 10 bolts (i) 2 (ii) none (iii) at most 2 bolts will be defective.

Solution. Probability of defective bolts, $p = 10\% = \frac{10}{100} = 0{\cdot}1$

$q = 1 - p = 1 - 0{\cdot}1 = 0{\cdot}9$

No. of bolts, $n = 10$

(i) Probability of 2 defective bolts,

$$P(2) = {}^{10}c_2 (\cdot 1)^2 (\cdot 9)^{10-2}$$

$$= 45 \times \cdot 01 \times (\cdot 9)^8$$

$$= 0{\cdot}45 \times (\cdot 9)^8$$

$$= 0{\cdot}1937$$

Also $P(1) = {}^{10}c_1 (\cdot 10)^1 \cdot (\cdot 9)^9 = 0{\cdot}38742$

(ii) Probability of none is defective;

$$P(0) = {}^{10}c_0 (\cdot 1)^0 (\cdot 9)^{10}$$

$$= 0{\cdot}3487$$

(iii) Probability of at most 2 defective bots

$$P(r \le 2) = P(0) + P(1) + P(2)$$
$$= 0{\cdot}3487 + 0{\cdot}38742 + {\cdot}1937$$
$$= 0{\cdot}92982$$

EXERCISE

1. An urn contains 5 white, 7 red and 8 blue balls. If 4 balls are drawn one by one with replacement what is the probability that

(i) none is white (ii) all are white

(iii) At least one is white (iv) Only 2 are white

Ans. $\frac{81}{256}, \frac{1}{256}, \frac{175}{256}, \frac{27}{256}$.

2. Six dice are thrown 729 times. How many times at least 3 dice will show a four or a five.

Ans. 233

3. 10 coins are tossed simultaneously. Find the probability of getting 7 heads.

$^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3$ **Ans.** $\frac{11}{64}$

4. If 20% of the bolts produced by a machine are defective, find he probability that out of 4 bolts chosen at random.

(i) 1 (ii) None (iii) at most 2 bolts will be defective

Ans. (i) 0·4096 (ii) 0·4096 (iii) 0·9728

5. Find the probability of getting a total of 7 at least once in 4 tosses of a pair of dice.

Ans. $\frac{671}{1296}$

6. Find the Bionomial Distribution whose mean in 5 and variance is $\frac{10}{3}$.

Ans. $^{15}C_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{15-r}$

7. A box contains 10 screws, 3 of which are defective. Two screws are drawn at random without replacement. Find the probability that none of the two screws is defective.

Ans. $\frac{7}{15}$

8. *The probability that a bulb manufactured in a factory will fuse after a use of 5 months is 0·05. Find the probability that out of 5 bulbs.*

(i) None (ii) At most one (iii) at least one will fuse after 5 months.

Ans. (i) $\left(\frac{19}{20}\right)^5$ (ii) $\frac{6}{5}\left(\frac{19}{20}\right)^4$ (iii) $1-\left(\frac{19}{20}\right)^5$

9. *If the probability of hitting a target is 10% and 10 shots are fired independently, what is the probability that the target will be hit at least once ?*

Ans. 0·649

10. *In a hurdle race, a player has to cross 10 hurdles. The probability that he will clear each hurdle is 5/6. Find the probability that he will knock down at least two hurdles.*

$p = 5/6$, $q = 1/6$

Ans. $\frac{8}{3}\left(\frac{5}{6}\right)^9$.

Poisson Distribution

Poisson's Distribution : Poisson Distribution is limiting case of Binomial Distribution when p is very small and n is very large i.e. $p \to 0$ and $n \to \infty$ such that $np = \lambda$, finite number. In Poisson's Distribution, probability, of r successes,

$$P(r) = \frac{\lambda^r \cdot e^{-\lambda}}{r!}, \text{ where } \lambda = np \text{ is mean of distribution.}$$

Proof. From Binomial Distribution, Probability of r successes,

$$P(r) = {}^nc_r p^r q^{n-r}$$

$$\Rightarrow P(r) = {}^nc_r\, p^r (1-p)^{n-r} \quad (\because p + q = 1)$$

$$\Rightarrow P(r) = \frac{n!}{r!(n-r)!}\left(\frac{\lambda}{n}\right)^r\left(1-\frac{\lambda}{n}\right)^{n-r}\left(\because np = \lambda \Rightarrow p = \frac{\lambda}{n}\right)$$

$$\Rightarrow P(r) = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!}\frac{\lambda^r}{n^r}\left(1-\frac{\lambda}{n}\right)^{n-r}$$

$$\Rightarrow P(r) = \frac{n(n-1)(n-2)\ldots(n-r+1)}{n^r}\cdot\frac{\lambda^r}{r!}\cdot\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$\Rightarrow P(r) = \frac{n(n-1)(n-2)\ldots(n-r+1)}{n\cdot n\cdot n_{(r\,times)}\cdot n}\cdot\frac{\lambda^r}{r!}\cdot\frac{\left[\left(1-\frac{\lambda}{n}\right)^{\frac{-n}{\lambda}}\right]^{-\lambda}}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$\Rightarrow P(r) = \lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{r-1}{n}\right)\frac{\lambda^r}{r!}\cdot\frac{\left[\left(1-\frac{\lambda}{n}\right)^{\frac{-n}{\lambda}}\right]^{-\lambda}}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$\Rightarrow P(r) = (1-0)(1-0)\ldots(1-0)\cdot\frac{\lambda^r}{r!}\cdot\frac{e^{-\lambda}}{(1-0)^r}, \text{ since } n \to \infty.$$

$$\Rightarrow P(r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

Mean of Poisson Distribution

$$P(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

We have, $\Sigma f = P(0) + P(1) + P(2) + \ldots + P(r) + \ldots$

$$= e^{-\lambda} + e^{-\lambda}\cdot\frac{\lambda}{1!} + e^{-\lambda}\cdot\frac{\lambda^2}{2!} + \ldots + \frac{e^{-\lambda}\lambda^r}{r!} + \ldots + \ldots$$

$$= e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\right]$$

$$= e^{-\lambda}\cdot e^{\lambda} = 1$$

$$\Sigma f.r = 0\cdot e^{-\lambda} + 1\cdot e^{-\lambda}\frac{\lambda}{1!} + 2\cdot e^{-\lambda}\cdot\frac{\lambda^2}{2!} + 3e^{-\lambda}\frac{\lambda^3}{3!} + \ldots +$$

$$= \lambda\cdot e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\right].$$

$$= \lambda\cdot e^{-\lambda}\cdot e^{\lambda}$$

$$\Sigma f\cdot r = \lambda$$

$$\therefore \text{ Mean} = \frac{\Sigma fr}{\Sigma f} = \frac{\lambda}{1} = \lambda$$

Mean $= np = \lambda$

Standard deviation $= \lambda + \lambda^2$

Standard Deviation of Poisson Distribution

We have, $\sigma^2 = \frac{\Sigma fr^2}{\Sigma f} - \left(\frac{\Sigma fr}{\Sigma f}\right)^2$

$$\sigma^2 = (m^2 + m) - (m)^2 = m$$

$$\Rightarrow \sigma = \sqrt{m} \quad [\because \Sigma fr^2 = m + m^2, \Sigma f = 1]$$

where $\Sigma fr^2 = \Sigma r^2\cdot P(r)$

$$= \Sigma r^2\cdot\frac{e^{-\lambda}\lambda^r}{r!}$$

$$= 0 + \frac{e^{-\lambda}\cdot\lambda}{1!} + 2^2\cdot\frac{e^{-\lambda}\cdot\lambda^2}{2!} + 3^2\cdot\frac{e^{-\lambda}\cdot\lambda^3}{3!} + \ldots$$

$$= e^{-\lambda}\cdot\lambda\left[1 + 2\cdot\lambda + \frac{3\lambda^2}{2!} + \frac{4\lambda^3}{3!} + \ldots\right].$$

$$\Rightarrow \Sigma fr^2 = e^{-\lambda}\cdot\lambda\left[\left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right) + \left(\lambda + 2\frac{\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \ldots\right)\right]$$

$$\Rightarrow \Sigma fr^2 = e^{-\lambda}\cdot\lambda\left[e^{\lambda} + \lambda\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\right)\right]$$

$$= e^{-\lambda}\cdot\lambda\,[e^{\lambda} + \lambda\, e^{\lambda}]$$

$$\Rightarrow \Sigma fr^2 = \lambda + \lambda^2$$

Recurrence Relation for Poisson Distribution

We have, by Poisson Distribution,

$$P(r) = \frac{e^{-\lambda}\cdot\lambda^r}{r!}.$$

$$P(r+1) = \frac{e^{-\lambda}\cdot\lambda^{r+1}}{(r+1)!}$$

$$\therefore \frac{P(r+1)}{P(r)} = \frac{e^{-\lambda}\cdot\lambda^{r+1}}{(r+1)!}\times\frac{r!}{e^{-\lambda}\cdot\lambda^r}$$

$$= \frac{\lambda}{r+1}$$

$$\Rightarrow P(r+1) = \frac{\lambda}{r+1}P(r)$$

Example 2. Five coins are tossed 3200 times. Find the probability of getting 5 heads 60 times using Poisson's Distribution.

Solution. Tossing one coin, probability of getting one head

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

$\therefore$ Probability of getting 5 heads from 5 coins,

$$p = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$\therefore$ Mean; $\lambda = np = 3200 \times \frac{1}{32} = 100.$

$\therefore$ Probability of getting 5 heads 60 times,

$$= \frac{e^{-\lambda}\lambda^{60}}{60!} = \frac{e^{-100}(100)^{60}}{60!}.$$

EXERCISE

1. A screw-making machine produces 2% defective screws and packs them in boxes of 500. What is the probability that a box contains 15 defective screws.

   Ans. 0·0347

2. If 2% bulbs are defective. What will be probability that at most 5 defective bulbs are found in a box of 200 bulbs.

   Ans. 0·785

3. A book of 585 pages contains 43 typographical errors. These errors are randomly distributed throughout the book. Find probability that 10 pages chosen at random, will be free from error.

   Ans. 0·4795

4. If 5% fuses are defective. If fuses are packed in boxes of 100 and not more than 10 fuses aer defective. Find the approximate probability that a box will not have the guaranteed quality.

   Ans. 0·01369

5. If the number of telephone calls on an operator received from 9·00 to 9·05 a.m. follow a Poission Distribution with mean 3. Find the probability that

   (a) the operator will receive no calls in the time interval tomorrow,

   (b) in the next 3 days the operator will receive only one call in the time interval

   Ans. (a) e^{-3} (b) $3 \times e^{-9}$.

np = 3 = λ mean

Normal Distribution

Normal Distribution : Normal Distribution is a continuous distribution which is derived from Binomial Distribution as particular case when number of trials n is very large and $p \to \frac{1}{2}$.

The probability density function of the normal distribution, f (x) is defined as

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

Where $-\infty \leq x \leq \infty$ and σ and μ are respectively standard deviation and mean of the distribution. The graph of the normal distribution is called normal curve which is bell-shaped and symmetrical about the mean μ. The two tails of the normal curve extend to $+\infty$ and $-\infty$ towards positive and negative directions of x-axis. These tails do not meet x-axis. The curve is unimodel and mean = mode. The line x = μ divides the area under normal curve above x-axis into two equal parts.

Hence mean = mode = median

The total area under the normal curve above x-axis is 1. The area under normal curve between $x = x_1$ & $x = x_2$ represents the probability between interval.

$$\text{The total area under curve} = \int_{-\infty}^{\infty} f(x)dx = 1.$$

Standard Normal Distribution

If x is a random variable with mean μ and standard deviation σ then the random variable $z = \frac{x-\mu}{\sigma}$ has normal distribution with mean 0 and standard deviation 1. This random variable z is called standard normal variable :

For standard normal distribution, the probability density function f (z) is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

which is free from any parameter. Normal distribution is limiting case of the Binomial Distribution. In this case $p = q = \frac{1}{2}$.

Area Under the Normal Curve :

$$\ln f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ by taking } z = \frac{x-\mu}{\sigma}, \text{ standard normal curve is formed.}$$

The total area under normal curve is 1. The area under normal curve is divided into two equal parts by line z = 0. The mean for the normal distribution is zero and standard deviation is 1.

Example 1. A sample of 50 C.F.L. bulbs were tested to get $\bar{x}$ = 12 months and s = 3 months. Supposing the data to be normally distributed find the probability that the C.F.L. has life

(i) more that 15 months.

(ii) less than 6 months.

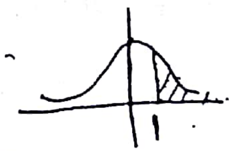
(iii) in between 9 and 15 months.

Solution. Let $z = \frac{x-\mu}{\sigma}$, given $\bar{x} = \mu = 12$

standard normal variable

(i) Here x = 15

$$\therefore z = \frac{x-\bar{x}}{\sigma} = \frac{15-12}{3} = 1$$



∴ Probability of life is more than 15 months,

$$P(x > 15) = P(z > 1)$$
$$= P(0 < z < \infty) - P(0 < z < 1)$$
$$= 0{\cdot}5 - 0{\cdot}3413$$
$$= 0{\cdot}1587$$

(ii) Here $x = 6$

$$z = \frac{x-\bar{x}}{\sigma} = \frac{6-12}{3} = -2$$

$\therefore$ $P(x < 6) = P(z < -2)$

$= P(z > 2)$

$= P(0 < z < \infty) - P(0 < z < 2)$

$= 0{\cdot}5 - 0{\cdot}4772$

$= 0{\cdot}0228$

$z < -2$
$z > 2$

(iii) Here $x_1 = 9$, $x_2 = 15$

$$\therefore z_1 = \frac{x_1-\bar{x}}{\sigma} = \frac{9-12}{3} = -1, \qquad z_2 = \frac{15-12}{3} = +1$$

$\therefore$ $P(9 < x < 15) = P(-1 < z < 1)$

$= P(-1 < z < 0) + P(0 < z < 1)$

$= P(0 < z < 1) + P(0 < z < 1)$

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

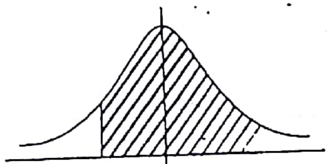
$= 2 \times 0{\cdot}3413$

$= 0{\cdot}6826$ *Answer.*

Example 2. The income of a group of 10000 persons was found to be randomly distributed with mean Rs. 750 p.m. and standard deviation Rs. 50. What percent of the group had income exceeding Rs. 668 and how many had income exceeding Rs. 832.

Solution. Given mean, $\mu = 750$, $\sigma = 50$

Standard normal variable, $z = \dfrac{x-\mu}{\sigma}$



for $x = 668$, $z = \dfrac{668-750}{50} = -1{\cdot}64$

$z = -1{\cdot}64$

$\therefore$ $P(x > 668) = P(z > -1{\cdot}64) = P(z < 1{\cdot}64)$

$= P(0 < z < \infty) + P(-1{\cdot}64 < z < 0)$

$= 0{\cdot}5 + P(0 < z < 1{\cdot}64)$

$= 0{\cdot}5 + 0{\cdot}4495$

$= 0{\cdot}9495$

$-1{\cdot}64$ 0 $1{\cdot}64$

$\therefore$ Percentage required $= 0{\cdot}9495 \times 100$

$= 94{\cdot}95\%$

Also for $x = 832$, $z = \dfrac{832-750}{50} = 1{\cdot}64$

$1{\cdot}64$

$\therefore$ $P(x > 832) = P(z > 1{\cdot}64)$

$= 0{\cdot}5 - P(0 < z < 1{\cdot}64)$

$= 0{\cdot}5 - {\cdot}4495$

$= 0{\cdot}0505$

$\cdot 5 - 1{\cdot}64$

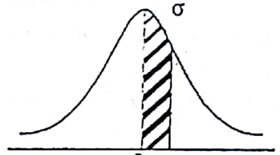
$\therefore$ No. of persons required $= 0{\cdot}0505 \times 10000$

$= 505$ ...wer. 505

TABLE - 1

AREA UNDER STANDARD NORMAL CURVE FROM $Z = 0$ TO $Z = \dfrac{x-\mu}{\sigma}$

An entry in the table is the proportion under the entire curve which is between $z = 0$ and a positive value of z. Area for negative values of z are obtained by symmetry.



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.8832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4415	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4841	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Scanned by CamScanner

EXERCISE

1. An aptitude test for employees in a bank is conducted on 1000 candidates. The mean store is 42 and standard deviation is 24. Under normal distribution find the no. of candidates whose scores exceed 60 and no. of candidates whose scores lie between 30 and 60.

   Ans. 227, 465.

2. In testing of 2000 bulbs, it was observed that mean $\bar{x}$ = 2040 hours and σ = 60 hours. Calculate how many bulbs have life
   (i) more than 2150 hrs.
   (ii) less than 1950 hrs.
   (iii) in between 1920 and 2180 hrs.

   Ans. (i) 67 (ii) 134 (iii) 1909

3. In aptitude test of 900 candidates, the mean score was 50 and S.D. = 20. Under normal distribution, find
   (i) No. of candidates whose mean score was less than 30.
   (ii) No. of candidates whose mean score was more than 70.
   (iii) No. of candidates whose mean score was in between 30 and 70.

   Ans. (i) 143 (ii) 143 (iii) 614

4. The mean height of 500 students is 151 cm. and the standard deviation is 15 cm. Assuming that height is normally distributed find how many student's height lie between 120 cm. and 155 cm.

   Ans. 294.

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

   Ans. mean = 50, S.D. = 10.

6. In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2·5. If the distribution is normal, find
   (i) how many student's score lies between 12 and 15.
   (ii) how many score above 18.
   (iii) how many score below 8.
   (iv) how many score 16.

   Ans. (i) 444 (ii) 55 (iii) 8 (iv) 116.

7. Find the mean and variance of the function $f(x) = \lambda e^{-\lambda x}$

   $xe^{-\lambda x}$

   Ans. $\frac{1}{\lambda}, \frac{1}{\lambda^2}$.

8. The life time of a certain component has mean life of 400 hours and standard deviation of 50 hours. Assuming the normal distribution for 1000 components find the number of components whose life time lies between 340 to 465 hours.

   Ans. 788.

9. The life of army shoses in normally distributed with mean 8 months and standard deviation 2 months. If 5000 pairs are issued, how many pairs would be expected to need replacement after 12 months ?

   Ans. 4886

Sampling Method

Sampling : Certain group of persons or individuals taken under study is called universe or population. A part of the universe (or population) is called sample which is finite subset of universe. The number of individual in a sample is called the size of sample. The process of selecting a sample from a universe is called sampling. The statistical constants of the population, such as mean (μ), standard deviation (σ) are called parameters.

The population parameters are not known generally. The sample characteristics are used to estimate the population. To estimate mean and standard deviation of the population is a primary purpose.

Sampling methodologies are of two categories first probability sampling and second non-probability sampling.

Probability samples allow the researcher to calculate the precision of the estimates obtained from the sample and to specify the sampling error.

Non-probability sample do not allow the studies finding to be generalized from the sample to the population.

Types of Sampling

Four types of sampling are :

1. Purposive Sampling
2. Random Sampling
3. Stratified Sampling
4. Systematic Sampling

Sampling Distribution of means from infinite population :

Let μ be mean and σ^2 be variance of population. Let x be random variable which denotes the measurement of the characteristic then.

Expected value of x, E (x) = μ

Variance of x, Var (x) = σ^2

The sample mean $\bar{x}$ is sum of n random variables $\frac{x_1}{n}, \frac{x_2}{n}, \ldots, \frac{x_n}{n}$.

Thus $E(x_1) = \mu, E(x_2) = \mu, \ldots, E(x_n) = \mu$

$Var(x_1) = \sigma^2, \ldots, Var(x_n) = \sigma^2$

$$E(\bar{x}) = E\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right]$$

$$= \frac{1}{n}E(x_1) + \frac{1}{n}E(x_2) + \ldots + \frac{1}{n}E(x_n)$$

$$= \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu$$

$$= \mu$$

$$Var(\bar{x}) = Var\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right]$$

$$= \frac{1}{n} var(x_1) + \ldots + \frac{1}{n} var(x_n)$$

$$= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \ldots + \frac{1}{n^2}\sigma^2$$

$$= \frac{\sigma^2}{n}$$

The expected value of the sample mean is the same as population mean but the variance of the sample mean is the variance of the population divided by n.

Sampling with Replecement :

When the sampling is done with replacement so that the population is back to the same form before the next sample member is picked up. Thus

$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \frac{\sigma^2}{n} \Rightarrow \frac{\sigma}{x} = \frac{\sigma}{\sqrt{n}}.$$

Sampling without Replacement from Finite Population :

when a sample is picked up without replacement from a finite population, the probability distribution of second variable depends on the outcome of the first picked up

$$E(x) = \mu,$$

$$Var(\bar{x}) = \frac{\sigma^2}{x} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

$$\Rightarrow \frac{\sigma}{x} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

If $\frac{n}{N} \to 0$ then

$$\frac{\sigma}{x} = \frac{\sigma}{n} \cdot \sqrt{\frac{1-\frac{n}{N}}{1-\frac{1}{N}}} \to \frac{\sigma}{\sqrt{n}}.$$

Sampling from Normal Population

If $x \sim N(\mu, \sigma^2)$ then $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Example 1. The diameter of a component produced on a semi-automatic machine is known to be normally distributed with mean of 10 mm and standard deviation 0·1 mm. If we pickup a random sample of size 5, what is probability that the same mean will be between 9·95 and 10·05 mm ?

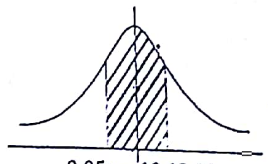
Solution. If the diameter of one component picked up at random be represent by variable x then.

$$x \sim N(10, 0\cdot1)$$

$$\Rightarrow \bar{x} \sim N\left(10, \frac{0\cdot1}{5}\right) \qquad \left[\because \bar{x} = N\left(\bar{x}, \frac{\sigma^2}{n}\right)\right]$$

$$P(r)\{9\cdot95 \le \bar{x} \le 10\cdot05\} = 2P(r)\{10 \le \bar{x} \le 10\cdot05\}$$

$$= 2.P(r)\left\{\frac{10-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{10\cdot05-\mu}{\frac{\sigma}{\sqrt{n}}}\right\}$$



$$= 2P(r)\left\{\frac{10-10}{\frac{0\cdot1}{\sqrt{5}}} \le \bar{z} \le \frac{10\cdot05-10}{\frac{0\cdot1}{\sqrt{5}}}\right\}$$

$$= 2P(r)\{0 \le \bar{z} \le \cdot5\sqrt{5}\}$$

$$= 2P(r)\{0 \le \bar{z} \le 1\cdot12\}$$

$$= 2 \times 0\cdot3686$$

$$= 0\cdot7372 \qquad \textit{Ans.}$$

Sampliing Distribution of the Variance

Sample variance is used to estimate the population variance. The sample variance s^2 is defined as $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$

Testing of Hypothesis and Null Hypothesis (Hco)

Certain decisions about population are taken on the basis of sample information. According to decision taken we make certain assumptions. Assuming the hypothesis correct we calculate the probability of getting the observed sample. If this probability is less then certain assigned value, then the hypothesis is rejected.

Null Hypothesis is the hypothesis if no difference and it is based on analysing the problem. First if is assumed that there is no difference between the observed value and expected value then it is tested whether the hypothesis is satisfied by the data or not. If the hypothesis is not approved the difference is taken significant. If the hypothesis is approved then difference is taken into account due to sampling function.

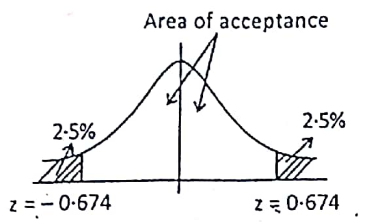
If Ho is rejected, the error is accepted. If Ho is accepted, the error is rejected.

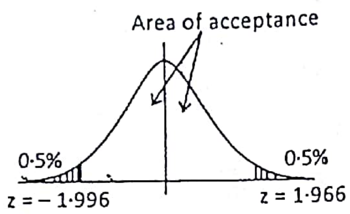
Level of Significance

The probability of the value of the variate falling in the critical region is called the level of significance. If the variate falls in the critical area, the hypothesis is rejected. The two critical regions which cover 5% and 1% of the normal curve are given below

Test of Significance

The test which enables us to decide whether to accept or reject the null hypothesis is called the test of significance. If the difference between the sample values and the population values lies in critical area then it is rejected.





Confidence limits

$\mu - 1{\cdot}96\sigma$, $\mu + 1{\cdot}96\sigma$ are 95% confidence limits as the area between $\mu - 1{\cdot}96\sigma$ and $\mu + 1{\cdot}96\sigma$ is 95%. If a sample statistics lies in the interval $\mu - 1{\cdot}96\sigma$ and $\mu + 1{\cdot}96\sigma$, we say 95% confidence interval.

$\mu - 2{\cdot}58\sigma$ and $\mu + 2{\cdot}58\sigma$ is 99% confidence limits.

Test of Significance of Large Samples

We know that Normal distribution is a limiting case of Binomial distribution as n is very larger. For normal distribution 5% of the items lie outside $\mu = \pm 1{\cdot}96\sigma$ while 1% items lie outside $\mu \pm 2{\cdot}58\sigma$.

The standard normal variate z and observed number of succes x are related by

$$z = \frac{x - \mu}{\sigma}.$$

After finding the value of z, test of significance is done depending upon the value of z.

1. If $|z| < 1{\cdot}96$ difference between observed and expected number of successes is not significant at the 5% level of significance.

   If $|z| > 1{\cdot}96$, difference is significant at 5% level of significance.

2. If $|z| < 2{\cdot}58$, the difference between observed and expected number of successes is not significant at the level of 1% of level of significance.

   If $|z| > 2{\cdot}58$, the difference is significant at 1% level of significance.

Example 2. A die was thrown 6000 times and 2 or 5 was obtained 2100 times. Can the deviation from expected value lie due to fluctuations of sampling ?

Solution. Let us consider the hypothesis that die is an unbiased one. Given $n = 6000$ the probability of getting 2 or 5 is

$$p = \frac{1+1}{6} = \frac{1}{3}$$

$$\therefore\ q = 1 - p = 1 - \frac{1}{3} = 2/3$$

$\therefore$ Expected value of number of successes

$$= np = 6000 \times \frac{1}{3} = 2000.$$

$$\sigma = \sqrt{npq} = \sqrt{6000 \times \frac{1}{3} \times \frac{2}{3}} = \sqrt{2000 \times \frac{2}{3}} = \sqrt{\frac{4000}{3}} = \sqrt{1333} = 11{\cdot}1$$

$$\therefore\ 3\sigma = 3 \times 11{\cdot}1 = 33{\cdot}3$$

Now, difference between actual number of successes and expected number of successes

$$= 2100 - 2000 = 100 \text{ which is greater than } 3\sigma = 33{\cdot}3.$$

Hence the hypothesis is incorrect and deviation is not due fluctuation of sampling.

Sampling Distribution of the Proportion

A sample of n items is drawn from population the probability of r success (r = 0, 1, 2, 3) are in terms of binomial expansion $(p + q)^n$.

In binomial distribution, mean = np, $\sigma = \sqrt{npq}$

1. Mean proportion of success $= \frac{\text{mean}}{n} = \frac{np}{n} = p$

2. Standard deviation of proportion of success $= \frac{\sigma}{n} = \frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$.

   It is also called standard error (S.E.)

3. Precision of the proportion of success $= \frac{1}{\text{S.E.}} = \frac{1}{\sqrt{\frac{pq}{n}}} = \sqrt{\frac{n}{pq}}$.

Example 3. A certain group reported 1700 sons and 1500 daughters. Does the data confirm the hypothesis that the sex ratio is $\frac{1}{2}$.

Solution. The total no. of observations = 1700 + 1500 = 3200

$\therefore$ The observed male ratio $= \frac{1700}{3200} = 0{\cdot}5313$

The male ratio according to given hypothesis = 0·5

$\therefore$ Difference between observed and theoretical ratio = 0·5313 – 0·5 = 0·0313

$\therefore$ The standard deviation of the proportion

$$s = \sqrt{\frac{pq}{n}} = \sqrt{\frac{{\cdot}5 \times {\cdot}5}{3200}} = 0{\cdot}0089$$

Since difference = 0·0313 is more than 3 × 0·0089 = 0·0267 therefore, the given data do not confirm the given hypothesis.

To Estimate Parameters of Population

The mean μ and standard deviation, σ are parameters of population whose estimates depend upon sample values. The mean and standard deviation of a sample are denoted by $\bar{x}$ and S.

An estimate of a population parameter given by single number is called a point estimation of parameter

$$(S.D.)^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

An interval in which population parameter lies with a given degree of confidence is called *interval* estimation.

(a) $(\bar{x}-\sigma, \bar{x}+\sigma)$, $\longrightarrow$ 68·27% confidence level

(b) $(\bar{x}-2\sigma, \bar{x}+2\sigma)$ $\longrightarrow$ 95·45% confidence level

(c) $(\bar{x}-3\sigma, \bar{x}+3\sigma)$ $\longrightarrow$ 99·37% confidence level

$\bar{x}$ and σ are mean and S.D. of sample.

Comparison of Large Samples

Let two large samples of size n_1 and n_2 be drawn from two populations of proportions of atributes p_1 and p_2 respectively.

Common proportion of atribute,

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

If e_1, e_2 be standard errors in two samples then

$$e_1^2 = \frac{pq}{n_1}, e_2^2 = \frac{pq}{n_2}$$

If e be standard error of the combined samples then

$$e^2 = e_1^2 + e_2^2 = \frac{pq}{n_1} + \frac{pq}{n_2} = pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right].$$

and $\quad z = \frac{p_1 - p_2}{e}$

Case I. If $z > 3$, the difference between P_1 and P_2 is significant

Case II. If $z < 2$, the difference is due to fluctuation of samples.

Case III. If $2 < z < 3$, the difference is significant at 5% level of significance.

Example 4. In a sample of 600 persons from a city, 450 are found smokers. In second sample of 900 persons from second city, 450 are smokers. Do the data indicate that the cities are significantly different w.r.t. the ratio of smoking among persons.

Solution. n_1 = 600, no. of smokers = 450

$$\therefore \; p_1 = \frac{450}{600} = 0{\cdot}75$$

n_2 = 900, no. of smokers = 450

$$\therefore \; p_2 = \frac{450}{900} = 0{\cdot}50$$

$$\therefore \; p = \frac{P_1n_1 + P_2n_2}{n_1 + n_2} = \frac{600 \times {\cdot}75 + 900 \times 0{\cdot}5}{600 + 900} = 0{\cdot}60$$

$$\therefore \; q = 1 - 0{\cdot}60 = 0{\cdot}40$$

$$\therefore \; e^2 = e_1^2 + e_2^2 = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

$$\Rightarrow e^2 = 0{\cdot}6 \times 0{\cdot}4\left(\frac{1}{600} + \frac{1}{900}\right)$$

$$\Rightarrow e^2 = 0{\cdot}000667$$

$$\Rightarrow e = 0{\cdot}02582$$

$$\Rightarrow z = \frac{p_1 - p_2}{e} = \frac{0{\cdot}75 - 0{\cdot}50}{0{\cdot}02582} = 9{\cdot}682$$

since, $z > 3$ therefore difference is significant.

The t-Distribution for Small Sample

The t-distribution is used to test the significance of

(i) the mean of a small sample

(ii) the difference between the means of two small samples or to compare two small samples.

(iii) the correlation coefficient let $\bar{x}$ be mean of $x_1, x_2, \ldots, x_n$ from random sample drawn from normal population with mean μ then

$$t = \frac{(\bar{x}-\mu)}{\frac{s}{\sqrt{n}}}, \text{ where } s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

Example 5. A machine which produces mica insulating washers for use in electric device to turn out washers having a thickness of a 10 mm. A sample of 10 washers have an average thickness 9·52 mm with a standard deviation of 0·6 mm. Find out t.

Solution. Given $\bar{x}$ = 9·52, μ = 10, s = 0·6, n = 10

$$\therefore \quad t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = -\frac{9{\cdot}52 - 10}{\frac{0{\cdot}6}{\sqrt{10}}} = -\frac{0{\cdot}48\sqrt{10}}{0{\cdot}6}$$

$$= -0{\cdot}8\sqrt{10} = -2{\cdot}53$$

Answer.

Example 6. The average life of sample of 100 C.F.L. bulbs is computed to be 1570 hours with S.D. = 120 hrs. The company claims that the mean life of the bulbs produced by it is 1600 hrs. Using the level of significance of 0·05, is the claim acceptable?

Solution. $\bar{x}$ = 1570, s = 120, n = 100, μ = 1600

$$\therefore \; t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{\sqrt{100}\cdot(1570-1600)}{120} = 2{\cdot}5$$

$\therefore$ t = 1·96 at the level of significance 0·05 and calculated value of t, 2·5 > 1·96 value)

$\therefore$ the claim is not acceptable.

Testing for Difference between Means of two small samples :

Let the mean and variance of first population be μ_1 and σ_1^2 and mean and variance of second population be μ_2 and σ_2^2.

If $\bar{x}_1$ be the mean of small sample of size n_1 from first population and $\bar{x}_2$ be the mean of small sample of size n_2 from second population

$$\text{then} \quad E(\bar{x}_1) = \mu_1, \; var(\bar{x}_1) = \frac{\sigma_1^2}{n_1}$$

$$E(\bar{x}_2) = \mu_2, \; var(\bar{x}_2) = \frac{\sigma_2^2}{n_2}$$

If the samples are independent, then $\bar{x}_1$ and $\bar{x}_2$ are also independent

$$\therefore \; E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

and

$$var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\therefore \; \bar{x}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \; \bar{x}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\therefore \; (\bar{x}_1 - \bar{x}_2) \sim N\left[\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]$$

$$\therefore \; t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

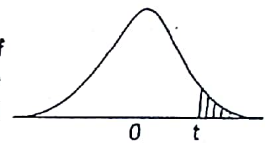
Case I. If $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2 = \sigma$ i.e. if population is same,

$$\text{then} \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} - \frac{1}{n_2}}}$$

TABLE - 2

STUDENTS' t DISTRIBUTION

The first column lists the number of degrees of freedom (v). The headings of the other columns give probabilities (P) for t to exceed the entry value. Use symmetry for negative t values.

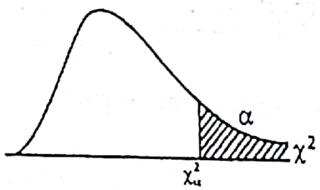


P/v	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.812	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.449
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.287
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
[illegible]	1.303	1.684	2.021	2.433	2.704
1	1.296	1.671	2.000	2.390	2.660
∞	1.289	1.658	1.980	2.358	2.617
[illegible]	1.282	1.645	1.960	2.326	2.576

TABLE - 3

CHI-SQUARE DISTRIBUTION

The following table provides the values of χ_α^2 that correspond to a given upper-tail area χ and a specified number of degrees of freedom.



Degree of Freedom	Upper-Tail Area					
	.20	.10	.05	.02	.01	.001
1	1.642	2.706	3.841	5.412	6.635	10.827
2	3.219	4.605	5.991	7.824	9.210	13.815
3	4.642	6.251	7.815	9.837	11.345	16.268
4	4.989	7.779	9.488	11.668	13.277	18.465
5	7.289	9.236	11.070	13.388	15.086	20.517
6	8.558	10.645	12.592	15.033	16.812	22.457
7	9.083	12.017	14.067	16.622	18.475	24.322
8	11.030	13.362	15.507	18.168	20.090	26.125
9	12.242	14.648	16.919	19.679	21.666	27.877
10	13.442	15.987	18.307	21.161	23.209	29.588
11	14.631	17.275	19.675	22.618	24.725	31.264
12	15.812	18.549	21.026	24.054	26.217	32.909
13	16.985	19.812	22.362	25.472	27.688	34.528
14	18.151	21.064	23.685	26.873	29.141	36.123
15	19.311	22.307	24.996	28.259	30.578	37.697
16	20.465	23.542	26.296	29.633	32.000	39.252
17	21.615	24.769	27.587	30.995	33.409	40.790
18	22.760	25.989	28.869	32.346	34.805	42.312
19	23.900	27.204	30.144	33.687	36.191	43.820
20	25.038	28.412	31.410	35.020	37.566	45.315
21	26.171	29.615	32.671	36.343	38.932	46.797
22	27.301	30.813	33.924	37.659	40.289	48.268
23	28.429	32.007	35.172	38.968	41.638	49.728
24	29.553	33.196	36.415	40.270	42.980	51.179
25	30.675	34.382	37.652	41.566	44.314	52.620
26	31.795	35.563	38.885	42.856	45.642	54.052
27	32.912	36.741	40.113	44.140	46.963	55.476
28	34.027	37.916	41.337	45.419	48.278	56.893
29	35.139	39.087	42.557	46.693	49.588	58.302
30	36.250	40.256	43.773	47.962	50.892	59.703

Example 1. Two independent samples of 8 and 7 items had respectively the following values of the variable.

Sample 1	9	11	13	11	15	9	12	14
Sample 2	10	12	10	14	9	8	10	

Is the difference between the means of the samples significant.

(given for V = 13, $t_{0.05} = 2{\cdot}16$)

Solution. $\bar{x} = \dfrac{\Sigma x}{n_1} = \dfrac{9+11+13+11+15+9+12+14}{8} = 11{\cdot}75$

Let assumed mean of x = 12

$$\bar{y} = \frac{\Sigma y}{n_2} = \frac{10+12+10+14+9+8+10}{7} = 10{\cdot}43$$

Let assumed mean for y is 10.

x	x − 12	$(x-12)^2$	y	y − 10	$(y-10)^2$
9	−3	9	10	0	0
11	−1	1	12	2	4
13	1	1	10	0	0
11	−1	1	14	−1	16
15	3	9	9	−2	1
9	−3	9	8	0	4
12	0	0	10		0
14	2	4			
94	−2	34	73	3	25

$$\sigma_x^2 = \frac{\Sigma(x-12)^2}{n_1} - \left\{\frac{\Sigma(x-12)}{n_1}\right\}^2$$

$$\sigma_x^2 = \frac{34}{8} - \left(\frac{-2}{8}\right)^2 = 4{\cdot}1875$$

$$\sigma_y^2 = \frac{\Sigma(y-10)^2}{n_2} - \left(\frac{\Sigma(y-10)}{n_2}\right)^2 = \frac{25}{7} - \left(\frac{3}{7}\right)^2 = 2{\cdot}13$$

$$\therefore\ s = \sqrt{\frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{34+25}{8+7-2}} = \sqrt{4{\cdot}54} = 2{\cdot}13$$

$\therefore\ s = 2{\cdot}13$

$$\therefore\ t = \frac{\bar{x}-\bar{y}}{s\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \frac{11\cdot75-10\cdot43}{2\cdot13\sqrt{\frac{1}{8}+\frac{1}{7}}}$$

$$\Rightarrow t = \frac{1\cdot32}{2\cdot13\times0\cdot518} = 1\cdot12$$

$\because\ t_{cal} = 1\cdot2 < t_{tab} = 2\cdot16$

i.e. calculated value of t is less then the tabular (given) value of t.

Therefore the difference between the means of samples is not siginificant.

EXERCISE

1. A die was thrown 96 times and show 2 upwards 84 times. Is the die biased ?

Ans. Yes.

2. Out of 800 children 300 are found to be under weight. Estimate, under simple sampling, the percentage of children who are underweight and also assign limits within which the percentage lies.

Ans. 37·5 %, limits = 37·5 ± 3 (2·4)

3. Out of 500 pens, 50 pens were found defective. Find percentage of defective pens and assign the limits within which the percentage probably lies.

Ans. 10%, 10 ± 3·9

4. In a certain city 20% of a random sample Of 900 students had eye defects. In second city, 18.5% of 1600 had the eye defect. Is the difference between the proportions significent ?

Ans. z = 0·37, the difference is significant

5. Two atheletes, Ram and Shaym where tested according to time (in seconds) to run a race. According to following data, test whether we can discriminate between two atheletes.

Athelete Ram	28	30	32	33	33	29	39
Shyam	19	30	30	24	27	29	

Ans. yes, with 75% confidence

6. 1000 bolts from a factory were examined and found to be 3% defective. 1500 another bolts from second factory were found to be 2% defective. Can it reasonably be concluded that the product of the first factory is inferior to second?

Ans. No, it cannot be concluded

Chi Square Test

Chi square distribution is the most popular distribution function in statistics. The square of a standard normal variate is called chi square (x^2) variate.

$z = \frac{x-\mu}{\sigma}$ is a normal variate white $\left(\frac{x-\mu}{\sigma}\right)^2$ is a chi square variate.

Let x be a normally distributed varlate and $x_1, x_2, ..., x_n$ be a random sample of n values then $w = x_1^2 + x_2^2 + ... + x_n^2$ has x^2 distribution with n degree of freedom.

Necessary Conditions

There are some necessary conditions for Chi square test.

1. The sample under study must be large. Total of cell frequency should not be less than 50.
2. The member of cells should be independent.
3. The cell frequency of each cell should be greater than 5.

Use of Chi-Square Test

It is used as

1. Test of independence
2. Test of goodness of fit
3. To test whether hypothesis value of population variate is σ^2.

Chi Square Test of Goodness of Fit

Chi-square test of goodness of fit is used to test the significance of the discrepancy between theory and experiment.

The theoretical frequencies of various classes are calculated from the assumption of the population. The significant deviation between the observed and theoretical frequencies is tested by using this test. χ is calculated from the formula

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

and $\Sigma O_i = \Sigma E_i = N$

Here O_i is the observed frequency and Ei is the expected or theoretical frequency of the cell.

The value of χ^2 is always positive such that $0 < \chi^2 < \infty$.

$\chi^2 = 0$ iff each pair is zero.

Degree of Freedom

1. For series of variables, Degree of freedom = (No. of items – 1)
2. When no. of frequency are put in cell, Degree of freedom = (R – 1) (C – 1)

Where R is no. of rows and C is no. of columns.

Example 1. A survey of 320 families with 5 children is given below :

No. of males	5	4	5	2	10	0
No. of Female	0	1	2	3	4	5
No. of Families	14	56	110	88	40	12

Is this result consistent with hypothesis ?

Solution. (1) Null Hypothesis (H_0)

Male and female both are equally probable

(2) Alternate Hypothesis (H_1)

Male and female both are not equally probable

Probability of male birth, $p = \frac{1}{2}$

Prob. of female birth, $q = \frac{1}{2}$

$$(p+q)^5 = \left(\frac{1}{2}\right)^5 + 5\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^4 + 10\left(\frac{1}{2}\right)^2\left(\frac{1}{\Sigma}\right)^3 + 10\left(\frac{1}{\Sigma}\right)^3\left(\frac{1}{\Sigma}\right)^2$$

$$+5\left(\frac{1}{2}\right)^4\left(\frac{1}{\Sigma}\right)+\left(\frac{1}{2}\right)^5 = \frac{1}{32}+\frac{5}{32}+\frac{10}{32}+\frac{10}{32}+\frac{5}{32}+\frac{1}{32}.$$

$$\therefore \text{ No. of females } = 320\left[\frac{1}{32}+\frac{5}{32}+\frac{10}{35}+\frac{10}{32}+\frac{10}{32}+\frac{5}{32}+\frac{1}{32}\right]$$

$$= 10 + 50 + 100 + 100 + 50 + 10$$

Expected frequencies of the female births.

O	E	O − E	$(O-E)^2$	$(O-E)^2/E$
14	10	4	16	1·6
56	50	6	36	0·72
110	100	10	100	1
88	100	− 12	144	1·44
40	50	− 10	100	2·00
12	10	2	4	0·40
			Total	7·16

Let level of significance, $\alpha = 0{\cdot}05$ critical value : The value of χ^2 at $\alpha = 0{\cdot}05$ for degree of freedom $(R-1)(C-1) = (6-1)(2-1) = 5$ is 11·07

$$\text{Calculate value of } \chi^2 = \frac{(O-E)^2}{E}$$

$$= 7{\cdot}16$$

$\because$ Calculate value of $\chi^2 = 7{\cdot}16$ is less then the tabular value of $\chi^2 = 11{\cdot}07$ at the level of significance 0·05 for degree of freedom 5.

$\therefore$ *the given result is consistent with the hypothesis Ho is male and female births are equally possible.*

Chi-square Test. As a Test of Independence :

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

Example 1. The I.Q. and economic condition of 1000 students are given below

Economic Con. \ I.Q.	High	Low	Total
Rich	100	300	400
Poor	350	250	600
Total	450	550	1000

Find whether there is any association between economic condition and I.Q. of students

Given for 1df, χ^2 at the level of significance 0·05 is 3·84.

Solution. H_0 :- There is no association between economic condition and I.Q.

H_1 :- There is association between E.C. and I.Q.

$$\text{Expected frequency, } E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

E.C. \ I.Q.	High	Low	Total
Rich	$\frac{400\times450}{1000} = 180$	$\frac{400\times550}{1000} = 220$	400
Poor	$\frac{600\times450}{1000} = 270$	$\frac{600\times550}{1000} = 330$	600
Total	450	550	1000

Calculation of χ^2 :-

O	E	O − E	(O − E)2	$(O-E)^2/E$
100	180	− 80	6400	35·5
350	270	80	6400	23·7
300	220	80	6400	29·1
250	330	− 80	6400	19.4

Degree of freedom, df = $(R-1)(C-1) = (2-1)(2-1) = 1$

Calculated $\chi^2 = 107{\cdot}7$

table value of $\chi^2 = 3{\cdot}84$

Calculated value of χ^2 is greater then the table value of χ^2

Here, H_0 is rejected, H_1 is accepted i.e. $\exists$ association b/w I.Q. and E.C.

EXERCISE

1. A die was thrown 90 times with following results.

Die face	1	2	3	4	5	6
Frequency	10	12	16	14	18	20

Test whether these data are consistent with hypothesis that the die is unbiased.

Given $\chi^2_{0{\cdot}05} = 11{:}07$ for 5 degree of freedom

Ans. Yes.

2. A survey amongst women was conducted to study the family life. The following observations were made :

Family life / Education	Happy	Unhappy	Total
Educated	70	30	100
Not educated	60	40	100
Total	130	70	200

Test whether there is any association between family life and education.

Ans. There is no association between family life and education

3. Records taken in the number of male and female births in 800 families are as below :

No. of male births	0	1	2	3	4	Total
No. of female births	4	3	2	1	0	
No. of families	32	178	290	236	64	800

Test whether the chance of a male birth is equal to that of a female birth.

Ans. Chances of male birth and female birth is equal.

4. 200 digits are chosen atrandom from a set of tables. The frequencies of the digits were

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

Use χ^2-test to assess the correctness of the hypothesis that digits were distributed in equal numbers in the table.

Given $\chi^2_{0.05}$ = 16·9, 18·3, 19·7 for 9, 10 and 11 degree of freedom respectively.

Ans. H_0 accepted.

5. The table below shows the results of drug against B.P.

	Not attacked	Attacked	Total
Drug	267	37	304
No drug	757	155	912
Total	1024	192	1216

Test whether there is any significant association between drug and attack.

Ans. Drug prevents the attack of B.P.

6. *Examine the goodness of fit binomial to the data given below :*

x	0	1	2	3	4	5	6	Total
f	6	20	28	12	8	6	0	80

Ans. The hypothesis that the data is taken from binomial population appears to be correct.

TIME SERIES

Definition : Time series is a kind of observations taken at time or space intervals arranged in chronological order. For example, production, population, prices and sales.

Uses

1. Time series is useful to know the past history of time series data.
2. It is used to predict future demand like production, prices, sales, weather conditions etc.
3. It helps in planning of the future operation.

Components of time Series

There are four types of components of Time series –

1. Secular Trends
2. Seasonal fluctuations
3. Cycliced fluctuations
4. Irregular component

Methods of Measuring Trends

1. Graphical method
2. Semi-average method
3. Moving-average method
4. Least square method

1. **Graphical Method :** This method is used in estimating secular trend. In this method a histogram is obtained by plotting the times series value on graph paper and then drawing a free hand smooth curve through these points. The curve should be smooth and number of points above and below should be equal. This method is simple and time saving method. It does not require any calculation.

2. **Method of Semi-average :** In this method, the data is divided into two equal parts and averages are computed of both the parts. In case of add number of years, the value of middle year is omitted. The average values of both the parts should be plotted against mid value of each half. By joining these two points we get the trend line which can be extended in both directions.

3. **Moving-average method :**

Case I. When period is odd

Let the period of moving average be 3 years with items a, b, c, d, e.

Average of First three = $\frac{a+b+c}{3}$ (It is written against b)

Average of second three = $\frac{b+c+d}{3}$ (It is written again c)

(dropping first and last)

Average of last three = $\frac{c+d+e}{3}$ (It is written against d)

Case II. When period is even

Let the period of moving average be four years then the average of first four figures will be placed between second and third year like wise the average of second group of four years will be replaced between third and fourth year. These two moving averages will then be averaged and this average will be written against the third year. This method is called the centering of the averages.

4. Method of least squares : Let y = a + bx be the line of best fit. Let x represents the time period and y represents the estimated value of the trend.

Let $E = \Sigma (y_i - y)^2$

$= \Sigma (y_i - a - bx)^2 = \Sigma (y - a - bx)^2$

For E to be minimum,

$$= \frac{\partial E}{\partial a} \ 0 \text{ and } \frac{\partial E}{\partial b} = 0$$

$\Rightarrow -2\Sigma (y - a - bx) = 0, -2\Sigma (y - a - bx)\, x = 0$

$\Rightarrow \Sigma y = xa + b\Sigma x$ ...(1)

$\Sigma yx = a\Sigma x + b\Sigma x^2$ ...(2)

Equations (1) and (2) are called normal equation of the line of best fit. Solving (1) and (2) for a and b and putting their values is y = a + bx we get equation of line of best fit.

To make calculation easy, we take the deviation from mean and putting $\Sigma x = 0$

We get, $\Sigma Y = na, \quad \Sigma XY = b\Sigma X^2$

where $X = x - \text{mean } \mu$,

$\therefore \quad a = \frac{\Sigma y}{n}, b = \frac{\Sigma xy}{\Sigma x^2}$.

We denote time ax x variable and trend values as y variable. First, we find the deviations of x from middle year then we find X^2 and XY. Putting the values of Σx, Σx^2, ΣXY *in the normal equations to get* $a = \frac{\Sigma y}{n}$ and $b = \frac{\Sigma xy}{\Sigma x^2}$.

At last we put the values of a and b in y = a + bx we get required line of best fit.

Example 1. *Find equation of line of best fit to the data given below :*

Year	2005	2006	2007	2008	2009	2010	2011
Population (in lakh)	30	35	55	64	72	80	83

Solution. *Suppose y = a + bX be the equation of line of best fit.*

year (x)	Population (in lac) y	X = x – 2008	X^2	XY
2005	30	– 3	9	– 90
2006	35	– 2	4	– 70
2007	55	– 1	1	– 55
2008	64	0	0	0
2009	72	1	1	72
2010	80	2	4	160
2011	83	3	9	249
	$\Sigma Y = 419$	$\Sigma X = 0$	$\Sigma X^2 = 28$	$\Sigma XY = 266$

$\therefore \quad a = \frac{\Sigma y}{n} = \frac{419}{7} = 59.86$

$b = \frac{\Sigma XY}{\Sigma x^2} = \frac{266}{28} = 9.5$

$\therefore \quad y = 59.86 + 9.5\,X$

Where X = x – 2008.

EXERCISE

1. Fit a trend line

Year	1971	1972	1973	1974	1975
Profit in lakhs	28	29·4	30·2	27	32·5

Ans. y = 28·7 + 0·35 (x – 1971·5)

2. Find the four yearly moving average from the data given below :

Year	1995	1996	1997	1998	1999	2000	2001	2002
Production (in quintal)	50	40	30	52	60	50	70	85

3. Find a trend line, using least square method to the data given :

Year	1975	1977	1979	1981	1983
Production (in quintal)	18	21	23	27	16

Ans. y = 21 + 0·2X

4. Fit a straight line trend by method of least squares to the following data :

Year	1979	1980	1981	1982	1983	1984	1985	1986
Earnings (in Rs.lakhs)	38	40	65	72	69	60	87	95

Ans. y = 65·75 + 3·66X

Statistical Quality Control Methods

Quality Control : Quality control is very important in case of product manufactured in a mill or factory. The sale and profit of the product depend upon the quality of a product. There are two ways to control the quality of product.

1. Physical Inspection 2. Statistical Quality Control

Types of Quality Control :

1. Process Control 2. Product Control

1. Process Control : The quality is controlled while the product is produced.

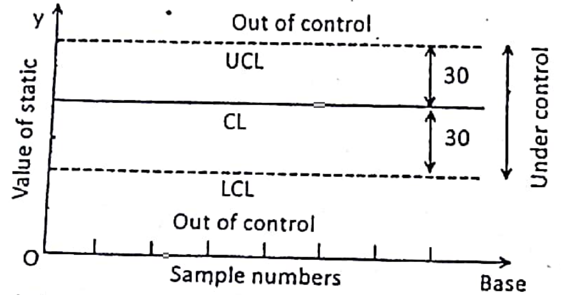
2. Product Control : The quality of product in checked before sale.

Control Chart : A graphical chart is actually a control chart which is used for presenting a sequence of sample characteristic. A chart defines the aim for the process. Chart is a means to maintain the goal and to judge whether the goal is achieved. Chart is a tool of statistical control.

In a control chart, there are three horizontal lines which are called control lines. There is vertical line which represents the quality statistic of every sample.

1. Central Line (C.L.) : Central line is in between middle of the chart and is parallel to the base line. Actually central line represents the prescribed standard quality of the product.

2. Upper Central Limit (UCL) : The upper control limit passes through the chart above and parallel to the central line. It represents the lower limit of the tolerance.



Types of Control Charts

1. Control charts for variable 2. Control charts for atributes

1. Control chart for variable : There are three types of control charts.

(a) Mean chart (b) Range chart (c) Standard deviation chart

(a) Mean chart : It is used for controlling quality of the product. A certain number of samples is taken at equal interval of time and mean calculated from each sample is used as statistic.

Let $x_1, x_2, x_3, ..., x_n$ be samples drawn from the process. Find mean of each sample

$$\bar{x} = \frac{\Sigma x}{n}.$$

Now, find the mean of sample means.

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}}{\text{No. of samples}}$$

At last find the standard error of means

$$\sigma\bar{x} = \frac{\bar{\sigma}}{\sqrt{n}}.$$

Then limits are

$$UCL = \bar{\bar{x}} + 3\sigma\bar{x}$$

$$LCL = \bar{\bar{x}} - 3\sigma\bar{x}$$

The control limits are :

$$UCL = \bar{\bar{x}} + A_2\bar{R}$$

$$LCL = \bar{\bar{x}} - A_2\bar{R}$$

∴ $\bar{R}$ is unbiased estimator of σ and value of A_2 can be obtained from table.

Example 1. Given number of samples is 20 and size of each sample is 5. If $\bar{R} = 2{\cdot}3\ \bar{\sigma}$, $\bar{\bar{x}} = 99{\cdot}5$. $\bar{R} = 7$. Find UCL and LCL for drawing a mean chart.

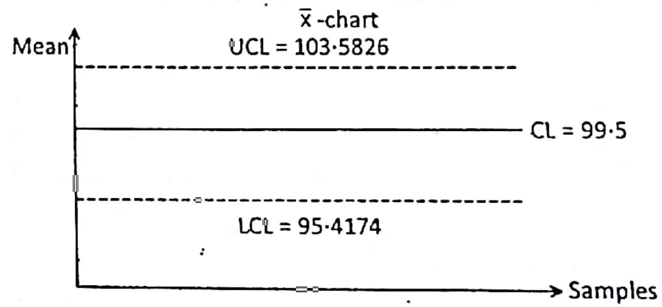
Solution. Given $\bar{\bar{x}} = 99{\cdot}5$, $\bar{R} = 7 = 2{\cdot}3\bar{\sigma}$, $n = 5$.

Now, $\bar{R} = 2{\cdot}3\bar{\sigma} \Rightarrow \bar{\sigma} = \frac{\bar{R}}{2{\cdot}3} = \frac{7}{2{\cdot}3} = 3{\cdot}043$

$$UCL = \bar{\bar{x}} + \frac{3\bar{\sigma}}{\sqrt{n}} = 99{\cdot}5 + \frac{3 \times 3{\cdot}043}{\sqrt{5}}$$
$$= 99{\cdot}5 + 4{\cdot}0826$$
$$= 103{\cdot}5826$$

$$LCL = \bar{\bar{x}} - \frac{3\bar{\sigma}}{\sqrt{n}} = 99{\cdot}5 - 4{\cdot}0826$$
$$= 95{\cdot}4174$$

CL = 99·5



(b) R-Chart

Construction of control chart for Range : Sample numbers are taken along x-axis and along y-axis, Range values are taken. We draw the control lines UCL & LCL. We plot the given points in the table to represent the range values.

With the help of R-chart, we can conclude whether the process variability is under control or not.

Example. The table given below gives the mean lengths and ranges of lengths of a product from 10 samples each of size 5. The limit for length are 200 ± 5cm. Construct $\bar{x}$-chart and R-chart and find whether the process is under control.

Sample No.	1	2	3	4	5	6	7	8	9	10
$\bar{x}$	201	198	202	200	203	204	199	196	199	200
R	5	0	7	3	3	7	2	8	5	5

Given for $n = 5$, $A_2 = 0{\cdot}58$, $D_4 = 2{\cdot}11$, $D_3 = 0$

Solution. Control limits for $\bar{x}$-chart

CL = 200, $\bar{\bar{x}} = \mu = 200$

$UCL_{\bar{x}} = \bar{\bar{x}} + A_2\bar{R}$

$LCL_{\bar{x}} = \bar{\bar{x}} - A_2\bar{R}$

Now, $\bar{R} = \dfrac{5+0+7+3+3+7+2+8+5+6}{10}$

$= \dfrac{46}{10} \simeq 4{\cdot}6$

$\therefore\ UCL_{\bar{x}} = \bar{\bar{x}} + A_2\bar{R}$

$= 200 + 0{\cdot}58 \times 4{\cdot}6 = 202{\cdot}69$

$LCL_{\bar{x}} = \bar{\bar{x}} - A_2R = 200 - 0{\cdot}58 \times 4{\cdot}6$

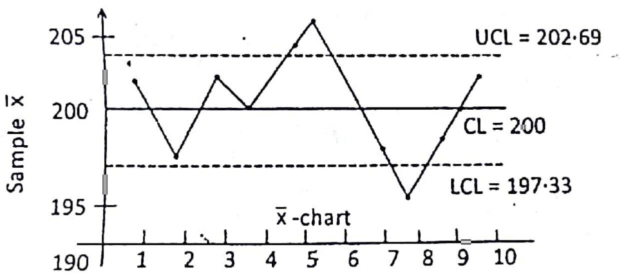
$= 197{\cdot}33$

Control limit for R-chart

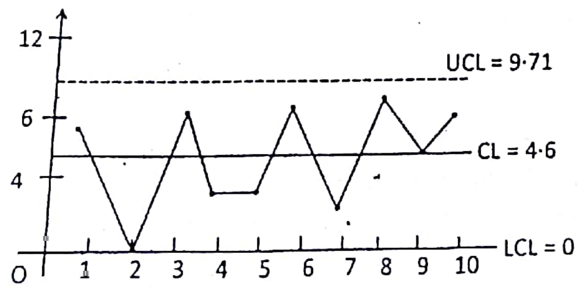
$CL = \bar{R} = 4{\cdot}6$

$UCL_{\bar{R}} = D_4\bar{R} = 2{\cdot}11 \times 4{\cdot}6 = 9{\cdot}71$

$LCL_R = D_3\bar{R} = 0 \times 4{\cdot}6 = 0$



$\bar{x}$-chart



R-chart

In R-chart, all points lie in the control limits therefore the process variability is under control. In X-chart, few points are outside the control limits, hence the process is not in statistical control.

EXERCISE

1. The product is packed by machines. 10 samples of size 5 were taken and the average weights were measured and the ranges were recorded below. Construct $\bar{x}$-chart and R-chart.

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean $\bar{x}$	15	16	17	18	14	15	15	17	17	18
Range R	4	7	7	7	5	8	9	11	12	14

Ans. $LCL_x = 20{\cdot}47$, $LCL_{\bar{x}} = 11{\cdot}93$, $UCL_R = 15{\cdot}65$, $LCL_R = 0$

2. Draw the mean and range charts.

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean	12·8	13·1	13·5	12·9	13·2	14·1	12·1	15·5	13·9	14·2
Range	2·1	3·1	3·9	2·1	1·9	3·0	2·5	2·8	2·5	2·0

given $n = 5$, $A_2 = 0{\cdot}577$, $D_3 = 0$, $D_4 = 2{\cdot}115$

Ans. CL = 5·4, UCL = 15·02, LCL = 12·04

3. Draw $\bar{x}$ and R-chart from the following chart.

Sample No.	1	2	3	4	5	6	7	8	9
$\bar{x}$	11	10·4	10·8	11·2	11·8	11·6	9·6	9·6	10
R	4	5	9	4	4	7	7	8	8

Ans. $UCL_{\bar{x}} = 14{\cdot}29$, $LCL_{\bar{x}} = 7{\cdot}03$, $UCL_R = 13{\cdot}32$, $LCL_R = 0$.

4. Draw the mean and range charts.

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean	5·10	4·98	5·02	3·96	4·96	5·05	4·94	4·92	4·92	4·98
Range	0·3	0·4	0·2	0·4	0·1	0·8	0·5	0·1	0·3	0·2

Given $A_2 = 0{\cdot}577$, $D_3 = 0$, $D_4 = 2{\cdot}115$, $n = 5$

Ans. $CL_{\bar{x}} = 4{\cdot}982$, $UCL_{\bar{x}} = 5{\cdot}19$, $LCL_{\bar{x}} = 4{\cdot}774$, $CL_R = 0{\cdot}36$, $UCL_R = 0{\cdot}761$, $LCL_R = 0$.

Fraction Defective Chart (P-chart) : P-chart is designed to control the proportion (p) percentage (100p)

Fraction defective (p) = $\dfrac{\text{No. of defective article}}{\text{size of sample}}$

$$p = \frac{\Sigma d}{N}$$

percentage defective = 100p

Mean Fraction Defective $\bar{p} = \dfrac{\text{No. of defectives in all samples}}{\text{To tal no. of items S in all the samples}}$

Rules to construct p-chart :

We calculate average fraction defective $\bar{p}$ then σ, the standard error of $\bar{p}$ is calculated.

$$\sigma = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Now, calculate UCL_p and LCL_p.

using $UCL_p = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

$$LCL_p = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Where $\bar{p}$ = mean defective of populations. Draw control line CL at $\bar{p}$ as thick horizontal line and two control lines UCL_p and LCL_p as dotted horizontal lines.

Then plot the individual points p. Interpretation :

(i) If all the points lie between two dotted lines then the method is called as statistical method.

(ii) If any point is above dotted line UCL_p or below the dotted line LCL_p then the process in called as out of statistical control. The control chart helps us to correct the faults. Warning can be given when the sample points lile out-side the control lines.

Example 1. 20 samples with size 210 each taken at the interval of one hour from a manufacturing process the average fraction defective was 0·067. Calculate the value of central line, upper and lower control line.

Solution. Average fraction defective = 0·067

$\therefore$ Central Line = $\bar{p}$ = 0·067

$$UCL_p = \bar{p} + 3\cdot\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= 0{\cdot}067 + 3\cdot\sqrt{\frac{0{\cdot}067(1-0{\cdot}067)}{210}}$$

$$= 0{\cdot}067 + 0{\cdot}0517 = 0{\cdot}1187$$

$$LCL_p = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0{\cdot}067 - 3 \times 0{\cdot}017 = 0{\cdot}01524$$ *Ans.*