23/11/23

PROBABILITY DISTRIBUTION AND SAMPLING

· CONTINUOUS PROBABILITY DISTRIBUTIONS:

Normal Distribution:

A random variable 'x' is said to be have a normal distribution is if its density function (or) probability distribution.

Probability Distribution:

It is given by $f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Here, μ = sigma standard Deviation $\left[\begin{matrix} -\infty < x < \infty \\ -\infty < \mu < \infty \end{matrix}\right]$

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

σ and μ be parameters.

Characteristics of Normal Curve:

- Normal distribution considers both +ve and -ve values.
- Normal distribution curve is symmetrical above ~~line~~ line $x = \mu$. It is in bell shaped curve.
- Area under the normal curve represents the total population.
- x-axis is asymtote to the curve.
- Here, mean = Median = Mode of the distribution coinsides at $x = \mu$.
- The total area bounded by the curve at x-axis is 1.

Mean of the Normal Distribution:

By the definition of Normal distribution,

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

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

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

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

Let $\frac{x-\mu}{\sigma} = z \Rightarrow x - \mu = z\sigma$

$x = z\sigma + \mu$

$dx = \sigma\, dz.$

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

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\, e^{-\frac{z^2}{2}} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz.$$

$$= 0 + \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-z^2/2} dz.$$

Let $\frac{z^2}{2} = t \Rightarrow z = \sqrt{2t}$

$z^2 = 2t$

$2z\, dz = 2dt$

$dz = \frac{dt}{\sqrt{2t}}$

$$= \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-t} \frac{dt}{\sqrt{2t}}$$

$$= \frac{2\mu}{\sqrt{2}\sqrt{2}\sqrt{\pi}} \int_0^\infty e^{-t} t^{\frac{-1}{2}} dt$$

$$\frac{\mu}{\sqrt{\pi}} \int_0^\infty e^{-t} t^{\frac{1}{2}-1} dt$$

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

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

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

$$= \frac{\mu}{\sqrt{\pi}} \Gamma\frac{1}{2} \qquad \left[\because \Gamma n = \int_0^\infty e^{-x} x^{n-1} dt\right]$$

$$= \frac{\mu}{\sqrt{\pi}} \sqrt{\pi} \qquad \left[\because \Gamma\frac{1}{2} = \sqrt{\pi}\right]$$

Mean $= \mu$

Variable of Normal Distribution :

By the definition of Normal Distribution :

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

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

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

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

Let $\frac{x-\mu}{\sigma} = z \Rightarrow$ $x = \sigma z + \mu$

$dx = \sigma\, dz$

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

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{\frac{-z^2}{2}} dz.$$

Let $\frac{z^2}{2} = t \Rightarrow z = \sqrt{2t}$

$$dz = \frac{dt}{\sqrt{2t}}$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} z^2 e^{-z^2/2} dz$$

Let $\frac{z^2}{2} = t$

$dz = \frac{dt}{\sqrt{2t}}$

$$= \frac{2 \times 2 \times \sigma^2}{\sqrt{2}\sqrt{2} \cdot \pi} \int_{0}^{\infty} e^{-t} \frac{t}{t^{1/2}} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-t} t^{3/2-1} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\frac{3}{2}$$

$$\left[\Gamma n = \int_{0}^{\infty} e^{-x} \cdot x^{n-1} \cdot dt\right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \Gamma\frac{1}{2}$$

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

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

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

$$= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \sqrt{\pi}$$

Variance $= \sigma^2$

12/12/2023

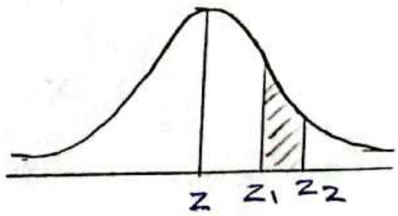
Procedure to find Probability density of normal curve:

The probability that the normal variable 'X' with mean 'μ' and ~~stand~~ standard deviation 'σ' lies between 2 special ~~varia~~ values x_1 and x_2 with $x_1 \le x_2$ can be obtained using area under the standard normal curve as follows.

Steps:

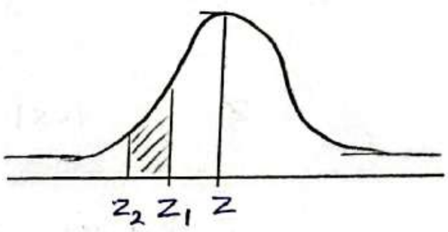
- Perform the change of scale $z = \frac{x-\mu}{\sigma}$ and find z_1, z_2 corresponding to the values x_1 and x_2 respectively.
- To find $P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$
- If both z_1, z_2 are positive then

$$P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$$
$$= |A(z_2) - A(z_1)|$$



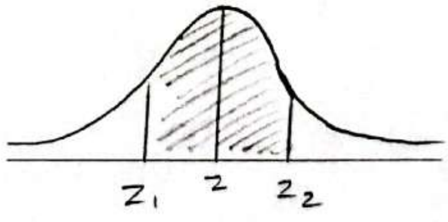
- If both z_1, z_2 are negative then,

$$P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$$
$$= |A(z_2) - A(z_1)|$$



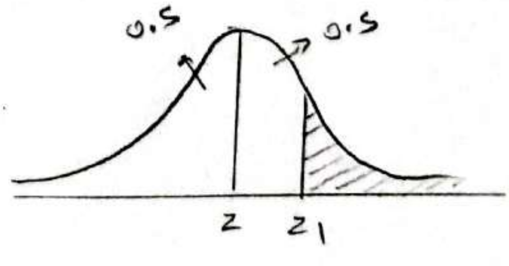
- If both $z_1 < 0$ and $z_2 > 0$ then

$$P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$$
$$= |A(z_1) + A(z_2)|$$



- If $z_1 > 0$ then

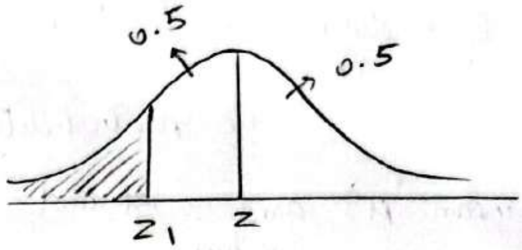
$$P(x > x_1) = P(z > z_1)$$
$$= 0.5 - (A(z_1))$$



- ~~If $z_1 < 0$ then~~
~~P(~~

• If $z_1 < 0$ then

$$P(x < x_1) = P(z < z_1)$$
$$= 0.5 - A(z_1)$$



14/12/2023

For a normally distributed variate with mean 1 and SD = 3.
Find the Probability that i) $P(3.43 \leq x \leq 6.19)$

ii) $P(-1.43 \leq x \leq 6.19)$

Given: $\mu = 1$

$\sigma = 3$

i) $P(3.43 \leq x \leq 6.19)$

Here, $x_1 = 3.43$, $x_2 = 6.19$

$$z_1 = \frac{X_1 - \mu}{\sigma} \qquad z_2 = \frac{x_2 - \mu}{\sigma}$$

$$= \frac{3.43 - 1}{3} \qquad = \frac{6.19 - 1}{3}$$

$$z_1 = 0.81 \qquad z_2 = 1.73$$

$z_1 > 0$ and $z_2 > 0$

$$P(x_1 \leq x \leq x_2) = P(z_1 \leq z \leq z_2)$$
$$= |A(z_2) - A(z_1)|$$
$$= |A(1.73) - A(0.81)|$$
$$= |0.4582 - 0.2910|$$
$$= 0.1672$$

H.W

Suppose the weights of 800 male students are normally destributed with mean 140 counts and SD 10 counts. Find the number of students whos weights are

i) between 138 - 148 counts 0.3674 // 294 ~

ii) more than 152 counts 0.115069 // 92.

15/12/23

In a normal distribution 7% of the items are under 35 and 89% under 63. Determine the mean of variance of the distribution.

Let μ = mean

σ = Standard Deveation.

Given: $P(x < 35) = 7\%$

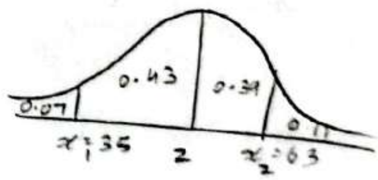
$P(x < 35) = 0.07$

$P(x < 63) = 89\%$

$= 0.89$

$P(x > 63) = 1 - 89\%$

$= 0.11$



$P(x < 35)$ $\qquad$ $P(x > 63)$

$x_1 = 35$ $\qquad$ $x_2 = 63$

$z_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{35 - \mu}{\sigma}$ $\qquad$ $z_2 = \dfrac{x_2 - \mu}{\sigma} = \dfrac{63 - \mu}{\sigma}$

$A(0.43) = \dfrac{35 - \mu}{\sigma}$ $\qquad$ $A(0.39) = \dfrac{63 - \mu}{\sigma}$

$1.48 = \dfrac{35 - \mu}{\sigma}$ $\qquad$ $1.23 = \dfrac{63 - \mu}{\sigma}$

$35 = 1.48(\sigma) + \mu$ $\qquad$ $63 = 1.23(\sigma) + \mu$

Solving, $\mu = 50$, $\sigma = 10.332$.

ii) $P(x < 64)$

$x_1 = \cancel{x}69$, $z_1 = \frac{x_1 - \mu}{\sigma}$

$= \frac{64 - 68}{3}$

$= \cancel{0.2}\ \cancel{0.33}\ \frac{-4}{3} = -1.33$

$P(x < 69) = P(z < 0.2)$

$= 0.5 - A(-1.33)$

$= 0.5 - 0.4082$

$= 0.0918$

No. of students with weight ~~g~~ less than or equal to 64 kg $= 0.0918 \times 300$

$= 27$ students.

iii) $P(65 \geq x \geq 71)$

$x_1 = 65$ $\qquad x_2 = 71$

$\cancel{x} \cdot z_1 = \frac{65 - 68}{3}$ $\qquad z_2 = \frac{71 - 68}{3}$

$z_1 = -1$ $\qquad z_2 = 1$

$P(65 \geq x \geq 71) = \cancel{x}\, P(-1 \geq z \geq 1)$

$= |A(1) + A(-1)|$

$= 0.6826$

No. of Students between 65 and 71 is 0.6826×300

$= 205$ students.

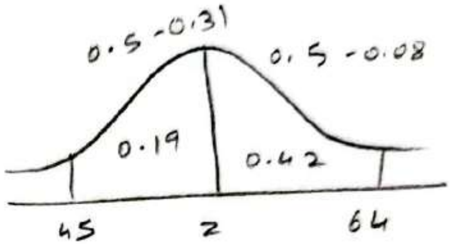
6

In a ND, 31% of the ~~the~~ items are under 45 and 8% are over 64. Find mean and variance of the destribution.

Let mean 'μ', variance 'σ' of ND.

$P(x < 45) = 0.31$ (31%)

$P(x > 64) = 0.08$ (8%)



$P(x < 45)$

Let $x_1 = 45$

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

$$z_1 = \frac{45 - \mu}{\sigma}$$

$0.5(\sigma) = 45 - \mu$

$P(x > 64)$

Let $x_2 = 64$

$$z_2 = \frac{x_2 - \mu}{\sigma}$$

$$z_2 = \frac{64 - \mu}{\sigma}$$

$1.41(\sigma) = 64 \mp \mu.$

$z_1 = A(0.19)$

$= 0.5$

$z_2 = A(0.42)$

$z_2 = 1.41$

Solving,

$\sigma = 9.84, \quad \mu = 50.$

$\sigma^2 = 97.$

H.W

Suppose 10% of the probability for a normal destribution is below 35 and 5% above 90. Find mean and standard deviation.

ii) $P(x \geq 45)$

$x_1 = 45 \qquad z_1 = \frac{45 - 30}{5}$

$z_1 = 3$

$P(x \geq x_1) = P(z \geq 3)$

$= 0.5 - A(z_1)$

$= 0.5 - 0.4987$

$= 0.0013.$

If the masses of 300 students are normally destributed with mean 68 kg and standard deveation 3 Kg.

How many students have masses i) greater than 72 Kg.

ii) less than or equal 64 kg.

iii) between 65 and 71 kg. Inclusive

$\mu = 68$ Kg

$\sigma = 3$ Kg

we have: i) $P(x > 72)$

$x_1 = 72, \quad z_1 = \frac{72 - 68}{3}$

$= 1.33$

$P(x > 72) = P(z > 1.33)$

$= 0.5 - A(z_1)$

$= 0.5 - A(1.33)$

$= 0.091211$

No. of Studuts weight greater than 72 is $P(x > 72) \times 300$

$= 0.0912 \times 300$

$= 27$ students.

ii) $P(-1.43 \le x \le 6.19)$

$x_1 = -1.43$, $x_2 = 6.19$

$z_1 = \frac{-1.43-1}{3}$ $\qquad z_2 = \frac{6.19-1}{3}$

$= -0.81$ $\qquad = 1.73$

$z_1 < 0$, $z_2 > 0$

$P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$

$= |A(z_1) + A(z_2)|$

$= |0.291 + 0.4582|$

$= 0.7492.$

If x is a normal variate with mean 30 and standard deviation 5. Find i) $P(26 \le x \le 40)$
ii) $P(x \geq 45)$

i) $P(26 \le x \le 40)$

$x_1 = 26$ $\quad x_2 = 40$

$z_1 = \frac{26-30}{5}$ $\quad z_2 = \frac{40-30}{5}$

$z_1 = -0.8 < 0$ $\quad = z_2 > 0$

$P(26 \le x \le 40) = |A(z_2) + A(z_1)|$

$= |A(2) + (A(0.8))|$

$= 0.4772 + 0.2881$

$= 0.7653.$

~~No. of students = 0.7653 × 3~~

16/12/23

The marks obtained in Math By 1000 students is normally distributed with mean 78% and ~~st~~ Standard deviation 11%. determine i) How many students got ~~abo~~ above 90%.

ii) What was the heighest marks obtained by the lowest 10% of the student.

iii) with in what limits did the middle of 90% of the students lie.

Solution: Mean = 78% = 0.78

S.D = 11% = 0.11

i) $P(X > 90\%) =$

Let $X_1 = 0.90$

$x_1 = 0.9$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{0.9 - 0.78}{0.11}$$

$z_1 = 1.09$

Hence the no. of students with marks more than 90%

$\Rightarrow P(X > 90\%) = \cancel{0.5} \cancel{P(z_1}$

$0.5 - A(z_1)$

$= 0.5 - 0.3623$

$= \cancel{0.113}\ 0.1377$

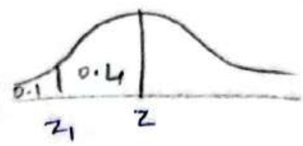
No. of Students $= P(X > 90\%) \times 1000$

$= 0.1377 \times 1000$

$= 138.$

ii)

$P(0.1) = 0.5 - \phi(0.4)$



$P(0.1) = 0.5 - A(-1.28)$

$z_1 = -1.281$

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

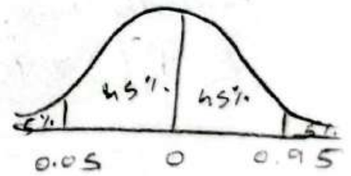
$$-1.281 = \frac{x_1 - 0.78}{0.11}$$

$x_1 = 0.6392$

Hence, Heighst mark obtained at botton 10% of the students is 63.92%

iii) Middle 90% corrosponding to 0.9 area leaving 0.05 on both sides. Then the corresprany z's are

$P(5\% \leq x \leq 95\%) = P(0.05 \leq z_1 \leq 0.95)$



$\therefore z_1 = -1.645$

$z_2 = +1.645$

$z_1 = -1.645$

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

$x_1 = 0.5985$

$z_2 = 1.645$

$$z_2 = \frac{x_2 - \mu}{\sigma}$$

$x_2 = 0.9615$

$\therefore$ Limits are are 59% to 96%.

(HW) In a sample of 1000 cases the mean of a certain test is 14 and Standard deveation 2.5. Assuming the destribution to be normal. Find i) How many students between 12 and 15.

19/12/2023

Normal Approximation to the binomial Destribution:

The normal destribution can be used to aproximate the binomial destribution. Suppose, the number of successors "X" ranges from X_1 to X_2. Then the probability of getting X_1 to X_2 successors is given by

$$\sum_{x=x_1}^{x_2} {}^nC_x \, p^x \, q^{n-x}.$$

- For ~~too~~ large "n", the calculation of binomial probabilities is very difficult. In such cases, we use normal destribution.

Note: For large "n" we can approximate, the binomial curve by the normal curve and calculate the probabilities.

- For any successors 'x', BL class interval is $(x-\frac{1}{2}, x+\frac{1}{2})$. Hence, z_1 must corrospond to the lower limit of the x_1 class and z_2 to the upperlimit of the x_2 class.

Hence, $$z_1 = \frac{(x_1 - \frac{1}{2}) - \mu}{\sigma} = \frac{(x_1 - \frac{1}{2}) - np}{\sqrt{npq}}$$

$$z_2 = \frac{(x_2 + \frac{1}{2}) - \mu}{\sigma} = \frac{(x_2 + \frac{1}{2}) - np}{\sqrt{npq}}$$

∴ Required probabilities:

$$P(x_1 \leq x \leq x_2) = P(z_1 \leq z \leq z_2) = \int_{z_1}^{z_2} \phi(z)\, dz$$

This can be evaluated using normal table.

Find the probability that out of 100 patences between 84 and 95 (inclusive) will survive a heart operation given that the changes (chanches) of survival is 0.9.

Given:

$p = 0.9$

$q = 1 - 0.9 = 0.1$

$n = 100$

$x_1 = 84, \quad x_2 = 95$

Mean of B.D $= np$

$= 100 \times 0.9$

$= 90.$

Variance of BD $= npq$

$= 100 \times 0.9 \times 0.1$

$\sigma^2 = 9$

Standard Deveation $= \sqrt{\sigma^2}$

$= \sqrt{9}$

$\sigma = 3$

$\therefore \; Z_1 = \dfrac{(x_1 - \frac{1}{2}) - \mu}{\sigma} \qquad Z_2 = \dfrac{(x_2 + \frac{1}{2}) - \mu}{\sigma}$

$= \dfrac{(84 - \frac{1}{2}) - 90}{3} \qquad Z_2 = \dfrac{(95 + \frac{1}{2}) - 90}{3}$

$Z_1 = -2.1\bar{6} \quad \left(\frac{-13}{6}\right) \qquad Z_2 = 1.8\bar{3} \quad \left(\frac{11}{6}\right)$

$\because \; P(x_1 \le x \le x_2) = \int_{z_1}^{z_2} \phi(z)\, dz = P(z_2 \le z \le z_1)$

$P(84 \le x \le 95) = \int_{-2.1\bar{6}}^{1.8\bar{3}} \phi(z)\, dz = P(-2.16 \le z \le 1.83)$

$\int_{-2.1\bar{6}}^{1.8\bar{3}} \phi(z)\, dz \therefore = P(-2.16 \le z \le 1.83)$

$= [A(1.83) + A(-2.16)]$

$= 0.46673 + 0.484869$

$= 0.9510$

8 coins are tossed together. Find the probability of getting 1-4 heads in a single toss.

$n = 8$

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

$x_1 = 1, \quad x_2 = 4$

Mean $= np$

$\mu = 8(\frac{1}{2})$

$= 4$

Variance $= npq$

$\sigma^2 = 8(\frac{1}{2})(\frac{1}{2})$

$= 2$

$SD = \sigma = \sqrt{2}$

$z_1 = \frac{(x_1 - \frac{1}{2}) - \mu}{\sigma}$ $\qquad z_2 = \frac{(x_2 + \frac{1}{2}) - \mu}{\sigma}$

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

$z_1 = -2.4748 \left(-\frac{7\sqrt{2}}{4}\right) \quad z_2 = 0.35355 \left(\frac{\sqrt{2}}{4}\right)$

$$= P(x_2 \le x \le x_1) = \int_{z_1}^{z_2} \phi(z)\, dz = P(z_2 \le z \le z_1)$$

$$\int_{z_1}^{z_2} \phi(z)\, dz = P(0.35355 \le z \le -2.4748)$$

$$= |A(z_2) + A(z_1)|$$

$$= A(0.35355) + A(+2.4748)$$

$$= 0.138161 + 0.49333$$

$$= 0.63149.$$

H.W

Find the probability that by guess work, a student can correctly answer 25- 30 questions. ~~Find the Probablity, by guss work~~ in a multiple choice quize consisting of 80 questions. Assume that in each question is with 4 choices only. Only one choice is correct and student has no knowledge of the subject.

$\mu = 20$
$\sigma = \sqrt{15}$
$z_1 = 1.1618$
$z_2 = 2.711$
$P(0.11428)$

SAMPLING DISTRIBUTION

Population: The number of observations under the study is known as population.

Eg: The number of electric bulbs manufacture in a company.

Size of Population: The number of observations in the population is defined to be the size of the population and it is denoted by "N".

Finite Population: The number of observations under the study is countable then it is known as finite population.

Eg: Number of workers in a factory.

Infinite Population: The number of observations under the study is uncountable then it is called infinite population.

Eg:- Number of stars in the sky.

Sample: Any finite subset of observations which are drawn from the population is known as sample. Sample is a part of the population.

Eg: Population: Total number of BTech colleges in India.

Sample: Total number of BTech colleges in Telangana.

Sample Space: The number of observations in sample is defined to be the size of the sample and it is denoted by "n".

Depending upon the sample size, Sample are classified into
i) Large Sample ii) Small Sample.

Large Sample: If $n \geq 30$ then it is called large sample.
Small Sample: If $n < 30$ then it is called small sample.

Parameter: The statistical measurement of the population is called parameter.

Eg: Population Mean (μ), population Variance (σ^2)

Statistic: Statistical measurement of the sample is called statistic.

Eg:- Sample Mean ($\bar{x}$), Sample variance (s^2).

Central Limit Theory: If $\bar{x}$ be the mean of the small size (n) drawn from the population with mean (μ) and SD (σ) then standadised sample $\bar{x} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

FORMULAS:

- Correction Factor $= \dfrac{N-n}{N-1}$
- Population Mean $(\mu) = \sum_{i=1}^{n} \dfrac{x_i}{N}$
- Population Variance $(\sigma^2) = \cancel{\sum_{i=1}^{n} \left(\frac{x_i-\mu}{N}\right)^2} \sum_{i=1}^{n} \dfrac{(x_i-\mu)^2}{N}$
- Sample Mean $(\bar{x}) = \sum_{i=1}^{n} \dfrac{x_i}{n}$
- Sample Variance $(s^2) = \sum_{i=1}^{n} \dfrac{(x_i-\bar{x})^2}{n-1}$

Infinite Population (with replacemt): (N^n)

- Mean of sample distribution of means:

$$\mu_{\bar{x}} = \frac{\text{Sum of all sample means}}{N^n}$$

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

- SD of the sample distribution of means

$$\sigma_{x} = \sqrt{\frac{\Sigma(x_i-\mu)^2}{N^n}}$$

Finite Population (without replacement): $({}^N C_r)$

- Mean of sample distribution:

$$\mu = \frac{\text{Sum of all sample mean}}{{}^N C_r}$$

- Variance of sample:

$$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{{}^N C_r}}$$

Sampling Distribution ①

Population : The no. of observations under the study is known as population.

Eg:- the no. of electric blubs manufactured in a company.

Size of population : the no. of observations in the population is defined to be the size of the population and it is denoted by N.

Finite population : the no. of observations under the study is countable then it is known as finite population.

Eg:- No. of workers in a factory.

Infinite population : The no. of observations under the study is uncountable then it is called infinite population

Eg:- No. of stars in the sky.

Sample: Any finite subset of observations which are drawn from the population is known as sample. Sample is part of a population.

Eg:- total no. of B.tech colleges in ~~Telangana~~ India is a population and tot no. of B.tech colleges in Telangana is a sample.

Sample space: The no. of observations in sample is defined to be the size of the sample and it is denoted by 'n'. depending up on the sample size. Sample has classified into 2 types.

(1) Large sample (2) Small sample.

Large sample: If $n \geq 30$ then it is called large sample.

Small sample: If $n < 30$ then it is called small sample.

parameter: The statistical measurement of the population is called parameter.

Eg:- population Mean (μ)
population variance (σ^2)

<u>Statistic</u> : Statistical measurement of the sample is called statistic.

Eg:- Sample Mean ($\bar{x}$)
Sample variance (s^2)

<u>Central limit Theorem</u> : If $\bar{x}$ be the mean of the small size (n) drawn from the population with mean (μ) and S.D (σ) then standardised sample $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

<u>formulas</u> :

(1) correction factor $= \dfrac{N-n}{N-1}$

(2) population Mean (μ) $= \sum_{i=1}^{n} \dfrac{x_i}{N}$

(3) population variance (σ^2) $= \sum_{i=1}^{n} \dfrac{(x_i-\mu)^2}{N}$

(4) sample mean ($\bar{x}$) $= \sum_{i=1}^{n} \dfrac{x_i}{n}$

(5) sample variance (s^2) $= \sum_{i=1}^{n} \dfrac{(x_i-\bar{x})^2}{n-1}$

Infinite population (with replacement) : (N^n)

(i) Mean of sample distribution of means

$$\mu_{\bar{x}} = \frac{\text{Sum of all sample mean}}{N^n}$$

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

(ii) S.D of the sample distribution of means

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N^n}}$$

finite population (without replacement) (NC_n)

(1) Mean of sample distribution

$$\mu = \frac{\text{Sum of all sample mean}}{NC_n}$$

(2) variance of sample

$$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{NC_n}}$$

(8) If N is size of population and n is sample size.

(1) If $n = 5$, $N = 200$ what is the value of correction factor

Sol :- correction factor $= \frac{N-n}{N-1} = \frac{200-5}{200-1} = \frac{195}{199} = 0.98$

(2) Find the value of the finite population correction factor for $n = 10$, $N = 1000$

Sol :- correction $= \frac{N-n}{N-1} = \frac{1000-10}{100-1} = \frac{990}{999} = 0.991$

(3) population consists of 5 numbers 2, 3, 6, 8, 11 consider all possible samples of size 2 which can be drawn with replacement from this population.

(i) Mean of population

(ii) S.D of population

(iii) The Mean of the sampling distribution of means and

(iv) S.D of sampling distribution of mean sol:- The total no. of samples with replacement

$$S = N^n = 5^2 = 25$$

(i) Mean of the population

$$\mu = \frac{\Sigma x_i}{N} = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

(ii) standard deviation of ~~sampling~~ population

$$\sigma = \sqrt{\frac{(x_i-\mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}}$$

$$\sigma = 3.29$$

(iii) The Mean of sampling distribution of means 2, 3, 6, 8, 11

$$\left.\begin{matrix} (2,2), & (2,3), & (2,6), & (2,8), & (2,11) \\ (3,2), & (3,3), & (3,6), & (3,8), & (3,11) \\ (6,2), & (6,3) & (6,6), & (6,8), & (6,11) \\ (8,2), & (8,3), & (8,6), & (8,8), & (8,11) \\ (11,2), & (11,3), & (11,6), & (11,8), & (11,11) \end{matrix}\right\} \text{sampling distributions}$$

(4)

The Mean of sampling distribution of means.

$$= \frac{2+2.5+4+5+6.5+2.5+3+4.5+5.5+7+4+4.5+6+7.0+8.5+5+5.5+7+8+9.5+7+8.5+9.5+11}{25}$$

$$= \frac{150}{25} = 6$$

(iv) S.D of sampling distribution of means

$$\sigma_{\bar{x}} = \sqrt{\frac{(\bar{x}-\mu)^2}{N^n}}$$

$$= \frac{(2-6)^2+(2.5-6)^2+(4-6)^2+\cdots+\cdots(9.5-6)^2+(11-6)^2}{25}$$

$$\sigma_{\bar{x}} = 2.32$$

—o—

~~Note~~

→ If the population is 3, 6, 9, 15, 27

(a) list all possible samples of size 3 that can be taken without Replacement from the finite population.

(b) calculate the mean of each of the sampling distribution of means.

(c) Find the S.D of sampling distribution of means.

Sol:- Mean of the population

$$\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$$

S.D of the population

$$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{n}}$$

$$= \sqrt{\frac{(3-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(27-12)^2}{5}}$$

$$= \sqrt{\frac{81+36+9+9+225}{5}} = \sqrt{\frac{366}{5}} = 8.4853$$

(a) Sampling distributions without replacement (finite population)

The total no. of samples with-out replacement

is $NC_n = 5C_3 = 10$

The 10 samples are

~~(3, 6, 9), (3, 6, 15), (3, 6, 27),~~ ~~(6, 9, 15), (6, 9, 27)~~

(9, 15, 27), ~~(3, 15, 27),~~ ~~(6, 15, 27), (3, 9, 15)~~, (3, 9, 27)

Scanned with ACE Scanner

computations

6 8 12, 10, 14, 16, 17, 9, 13

6 8 9 10 12 13 14 15 17

The 10 samples are

3, 6, 9, 15, 27

(3, 6, 9), (3, 6, 15), (3, 6, 27)
(6, 9, 15) (6, 9, 27), (9, 15, 27)
(3, 9, 15) (3, 9, 27), (3, 15, 27)
(6, 15, 27)

$\left[\frac{3+6+9}{3} = \frac{18}{3} = 6 \right]$

6, 8, 12, 10, 14, 17, 9, 13, 15, 16

computations

	6,	8	12	10	14	17	9	13	15	16
	6	8	9	10	12	13	14	15	16	17
x_i	6	8	9	10	12	13	14	15	16	17
f_i	1	1	1	1	1	1	1	1	1	1

(B) Mean of sampling distribution of sampling means is

$$\mu = \frac{6+8+9+10+12+13+14+15+16+17}{10} = \frac{120}{10} = 12$$

(c) S-D of sampling distribution of means

$$\sigma^2 = \frac{(6-12)^2+(8-12)^2+ \cdots (16-12)^2+(17-12)^2}{10}$$

$$\sigma^2 = \frac{120}{10} = 12$$

$$\sigma = 3.4$$

problem (3) solve without - replacement

sol :- (a) $\mu = 6$

(b) $\sigma = 3.29$

(c) sampling without replacement

the tot no. of samples without replacement $NC_n = 5C_2 = 10$ samples of size 2

2, 3, 6, 8, 11

The samples are

$$\left\{ \begin{matrix} (2,3) & (2,6) & (2,8) & (2,11) \\ (3,6) & (3,8) & (3,11) & \\ (6,8) & (6,11) & & \\ & (8,11) & & \end{matrix} \right\}$$

The selection (2,3) is considered same as (3,2).

The corresponding sampling means are

2.5 4 5 6.5

4.5 5.5 7

7 8.5

9.5

The mean of sampling distribution of means

is

$$\mu = \frac{2.5+4+5\ 7 -- + 8.5+9.5}{10} = \frac{60}{10} = 6$$

(d) The variance of sampling distribution of means.

$$\sigma^2 = \frac{(2.5-6)^2 + (4-6)^2 + -- + (8.5-6)^2 + (9.5-6)^2}{10}$$

$\sigma^2 = 4.05$

$\sigma = 2.01$

H.W

(1) let $S = \{1, 5, 6, 8\}$. find the probability distribution of the sampling mean for random sample of size 2 drawn without replacement. Ans: $\bar{x} = 5$

$\sigma = 1.612$

(3) samples of size 2 are taken from the population 1, 2, 3, 4, 5, 6

(i) with Replacement (ii) without replacement

find (a) The mean of the population

(b) standard deviation of population

(c) The mean of the sampling distribution of means.

(d) The standard deviation of the sampling distribution of means. 2

Sol:- (i) with Replacement

(a) The Mean of the population

$$\mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

(b) The variance of the population

$$\sigma^2 = \sum \frac{(x_i - \mu)^2}{N}$$

$$= \frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6}$$

$$= \frac{6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25}{6}$$

$$\sigma^2 = \frac{17.50}{6} = 2.917$$

S.D of the population is $\sigma = \sqrt{2.917}$

$\sigma = 1.71$

(c) Number of samples of size 2 with Replacement is $N^n = 6^2 = 36$

$$\left\{ \begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix} \right\} = 36$$

the no. of samples $n = 36$

Their means are

$$\left\{ \begin{matrix} 1 & 1.5 & 2 & 2.5 & 3 & 3.5 \\ 1.5 & 2 & 2.5 & 3 & 3.5 & 4 \\ 2 & 2.5 & 3 & 3.5 & 4 & 4.5 \\ 2.5 & 3 & 3.5 & 4 & 4.5 & 5 \\ 3 & 3.5 & 4 & 4.5 & 5 & 5.5 \\ 3.5 & 4 & 4.5 & 5 & 5.5 & 6 \end{matrix} \right\}$$

The mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{1 + 1.5 + 2 + \cdots + 5.5 + 6}{36}$$

$$\mu = \frac{126}{36} = 3.5$$

$$\boxed{\mu = 3.5}$$

(d) The Standard deviation of sampling distribution of means is

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$\sigma^2 = \frac{(1-3.5)^2 + (1.5-3.5)^2 + (2-3.5)^2 + \cdots + (5.5-3.5)^2 + (6-3.5)^2}{36}$$

$$\sigma^2 = \frac{52.5}{36}$$

$$\sigma^2 = 1.46$$

$$\sigma = 1.20$$

(ii) <u>without Replacement</u>

(a) The Mean of the population

$$\mu = 3.5$$

(b) the variance of the population

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = 2.917$$

$$\sigma^2 = 2.917$$

Number of samples of size 2 without replacement = NC_n , $N=6, n=2$

$$6C_2 = \frac{6!}{2!4!} = 15$$

They are
$$\begin{Bmatrix} (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,3) & (2,4) & (2,5) & (2,6) & \\ (3,4) & (3,5) & (3,6) & & \\ (4,5) & (4,6) & & & \\ (5,6) & & & & \end{Bmatrix}$$

corresponding means are
$$\begin{Bmatrix} 1.5 & 2 & 2.5 & 3 & 3.5 \\ 2.5 & 3 & 3.5 & 4 & \\ 3.5 & 4 & 4.5 & & \\ 4.5 & 5 & & & \\ 5.5 & & & & \end{Bmatrix}$$

15 samples

(c) Mean of sampling distribution of means

$$\mu = \frac{1.5+2+2.5+3+3.5+ \cdots +5.5}{15}$$

$$\mu = \frac{52.5}{15}$$

$$\mu = 3.5$$

(d) The standard deviation of sampling distribution of means is

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$\sigma^2 = \frac{(1.5-3.5)^2 + (2-3.5)^2 + (2.5-3.5) + \dots + (3.5-3.5)^2}{15}$$

$$\sigma_{\bar{x}}^2 = \frac{17.5}{15} = 1.16$$

$$\sigma = \sqrt{1.16}$$

$$\sigma = 1.077$$

17-6

① Find the Mean and S.D of sampling distribution of variancey for the population 2, 3, 4, 5 by drawing samples of size 2

(a) with Replacement

(b) without Replacement (finite population)

standard Errors

formulas :-

1. standard error of sample Mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$
2. S.E of sample proportion $p = \sqrt{\frac{PQ}{n}}$

   where $Q = 1 - P$
3. S.E of sample S.D $(s) = \frac{\sigma}{\sqrt{2n}}$
4. S.E of difference of two sample means $\bar{x_1}$ and $\bar{x_2}$ P/S

   $$(\bar{x_1} - \bar{x_2}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

   where n_1 and n_2 are sample sizes drawn from population.

   σ_1 and σ_2 are standard deviation respectively.
5. when samples are drawn from same population then S.E of difference of two samples means $\bar{x_1}$ and $\bar{x_2}$ $= \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

(6) when samples are having S.D s_1, s_2 for 2 samples whose means are $\bar{x}_1, \bar{x}_2$ then S.E of 2 samples means

$$\bar{x}_1 \& \bar{x}_2 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(or) $\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

where $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2}$

(7) S.E of difference of 2 sample proportions p_1, p_2

$= \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$ where P_1, P_2 are population proportions.

$= \sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ when samples are come from same population.

$= \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n}}$ when population proportions are not given.

$= \sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ where $p = \frac{n_1 P_1 + n_2 P_2}{n_1+n_2}$

$Q = 1-p.$

Note:

for a finite population of size N, when a sample is drawn without Replacement we have

(i) S.E of sample Mean $= \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

(ii) S.E of sample proportion $= \sqrt{\frac{PQ}{n}} \sqrt{\frac{N-n}{N-1}}$

* For an infinite population when the sample is drawn without replacement formula 1 and 2 remain the same

Here $\frac{N-n}{N-1}$ is correction factor.

(1) The Mean height of students in a college is 155 cms and standard deviation is 15. what is the probability that the mean height of 36 students is less than 157 cms.

Sol :- $\mu =$ Mean of the population

$\mu =$ Mean height of students of a college $= 155$ cm.

$\sigma =$ S.D $= 15$ cms

$n =$ sample size $= 36$

$\bar{x}$ = Mean of a sample = 157 cms

Now $z = \frac{x-\mu}{\sigma/\sqrt{n}} = \frac{157-155}{15/\sqrt{36}} = \frac{2}{15/6} = \frac{12}{15}$

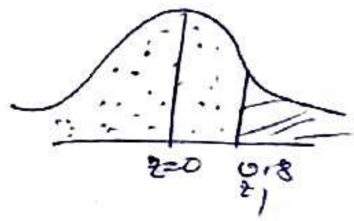
$z = 0.8 > 0$

$P(x \leq 157) = P(z < 0.8)$

$= 0.5 + A(0.8)$

$= 0.5 + 0.2881$

$= 0.7881$



thus the probability that the mean height of 36 students is less than 157 = 0.7881

② A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and its variance $\sigma^2 = 256$ what is the probability that $\bar{x}$ will be between 75 and 78.

Estimation :

Estimate : An estimate is a statement made to find an unknown population parameters is called estimate

σ, μ be the population parameters.

Estimator : The procedure to determine an unknown population parameters is called as estimator.

Types of estimation

we have 2 types of estimation

1. point estimation.
2. Interval estimation.

point estimation : If an estimation of the population is given by a single value then the estimate is called a point estimatio

Eg:- If the height of the student is 162 cm then The measurement gives a point estimation.

Interval estimation : If an estimation of the population is given by two different values between which the parameter lies then the estimate is called interval estimation.

eg:- If the height of the student is given by 165 ± 3.5 cm

confidence Interval. If $\bar{x}$ be the mean of the random sample size n from the population with variance σ^2, 100 % confidence interval for μ is given by

$$\left(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right),$$ Here α is called the level of significance.

Maximum error for large sample :

1. Max error is given by

$$E_{max} = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

2. If σ is not given

$$E_{max} = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

Maximum error for small sample :

$$E_{max} = t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

where $t_{\alpha/2}$ is the t- distribution with $v = (n-1)$ degrees of freedom.

→ for sample $E_{max} = z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$

→ confidence interval $\left[\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right]$.

V.Imp

Confidence limits :

1. $z_{\alpha/2} = 1.96$ for 5% level of significance (95% confidence)
2. $z_{\alpha/2} = 1.645$ " 10% " " " (90% confidence)
3. $z_{\alpha/2} = 2.58$ " 1% " " " (99% ")
4. $z_{\alpha/2} = 2.33$ " 2% " " " (98% ")

problems

1. A random sample of size 100 has a standard deviation of 5. what can you say about the maximum error with 95% confidence

Sol :- Given $n = 100$, S.D $\sigma = 5$

$z_{\alpha/2}$ for 95% confidence $= 1.96$

we know that maximum error

$$E_{max} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$= (1.96) \frac{5}{\sqrt{100}}$$

$$= 0.98$$

(2) The mean and standard deviation of a population are 11.795 and 14.054 respectively what can one assert with 95% confidence about the maximum error if $\bar{x} = 11.795$ and $n = 50$. And also construct 95% confidence interval for the true mean.

(or)

The Mean and standard deviation of a population are 11.795 and 14.054 respectively. If $n = 50$ find 95% confidence interval for the mean.

Sol :- Mean of population $\mu = 11.795$

S.D of population $\sigma = 14.054$

$\bar{x} = 11.795$,

n = sample size $= 50$

Maximum error $E_{max} = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$= 1.96 \frac{(14.054)}{\sqrt{50}}$

$= 3.8955$

$\therefore$ confidence interval $= \left(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$

$= (11.795 \pm 3.8955)$

$= (11.795 + 3.8955,\ 11.795 - 3.8955)$

$= [15.6905, 7.8994]$ as.

$\therefore$ confidence interval $(15.6905,\ 7.8994)$

(3) A random sample of size 100 is taken from a population with $\sigma = 5.1$ Given that the sample mean is $\bar{x} = 21.6$ construct a 95% confidence interval for the population mean ie.

sol:- Given $\bar{x}$ = Sample mean $= 21.6$

$z_{\alpha/2} = 1.96$,

n = sample size $= 100$, S.D $\sigma = 5.1$

$\therefore$ confidence interval $\left(\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$

$$= \left(\bar{x} \pm \frac{1.96 \times 5.1}{\sqrt{100}}\right)$$

$$= \cancel{21.6 \pm 1.9}$$

$$= \left(21.6 \pm \frac{1.96 \times 5.1}{10}\right)$$

$$= (21.6 \pm 0.999)$$

$$= (21.6 + 0.999,\ 21.6 - 0.999)$$

$$= (22.599,\ 20.601)$$

Hence $(20.601, 22.599)$ is the confidence interval for the population mean μ.

(4) what is the maximum error one can expect to make with probability 0.90 when using the mean of a random sample of size $n=64$ estimate the mean of population with $\sigma^2 = 2.56$.

Sol: Here $n = 64$

The probability $= 0.90$

$\sigma^2 = 2.56 \Rightarrow \sigma = \sqrt{2.56} = 1.6$

confidence limit 90%.

$z_{\alpha/2} = 1.645$

Maximum error $E_{max} = z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

$$= 1.645 \left(\frac{1.6}{\sqrt{64}}\right)$$

$$E_{max} = 0.329.$$

(5) If we can assert with 95% that the maximum error is 0.05 and $p = 0.2$. find the size of the sample.

Sol:- Given $p = 0.2$, $E_{max} = 0.05$

$$Q = 1 - p$$
$$= 1 - 0.2 \qquad z_{\alpha/2} = 1.96 \text{ (for 95\% confidence)}$$
$$= 0.8$$

we know that $E_{max} = z_{\alpha/2} \sqrt{\frac{PQ}{n}}$

$$0.05 = 1.96 \sqrt{\frac{0.2 \times 0.8}{n}}$$

squaring on both sides.

$$n = \frac{(1.96)^2 (0.2 \times 0.8)}{(0.05)^2}$$

$$n = 246.$$

Imp

(6) Find 95% confidence limits for the mean of a normality distributed population from which the following sample was taken.

15, 17, 10, 18, 16, 9, 7, 11, 13, 14

Sol:- we have $\bar{x} = \frac{15+17+10+18+16+9+7+11+13+14}{10}$

$n = 10$ $\qquad = \frac{130}{10}$

$$\bar{x} = 13$$

s^2 = Mean of sampling distribution of Means

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

$$= \frac{(15-13)^2 + (17-13)^2 + (10-13)^2 + \cdots + (13-13)^2 + (14-13)^2}{10-1}$$

$$= \frac{40}{10-1}$$

$$s^2 = 40/9$$

$$s = \sqrt{40/3}$$

for 95% confidence $z_{\alpha/2} = 1.96$

confidence limits are $= \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

$$= 13 \pm 1.96 \left(\frac{40}{3}\right)$$

$$= (13 \pm 2.26) \sqrt{10}$$

$$= (13 + 2.26,\ 13 - 2.26)$$

$$= (15.26,\ 10.74)$$

=

UNIT-III Chapter-2 Remaing Topics

t-Distribution (OR) students t-distribution - Small Samples

Introduction:- when the size of the sample (n) less than 30, then that sample is called a small sample, ie $n<30$, the sample is called a small sample. Sampling distribution approches a normal distribution and values of simple statistic are considered best estimates of the parameters in a population in case of large samples we can use normal distribution to test for a specified population mean (or) equality of means, in small samples only if the sample is drawn from a normal population whose standard deviation is known it involves the concept of "degrees of freedom".

→ The different types of distributions that involve sized samples where $(n<30)$ are

i) t-distribution ii) f-distribution.

Degree's freedom:- The number of degrees of freedom is equal to the total number of observations less than the number of independent constraints imposed on the observations. It is denoted by ν.

Def:- t-distribution:-

If $x_1, x_2, x_3 \cdots x_n$ be any random sample of size n drawn from a normal population with mean μ and variance σ^2, then the test statistic t is defined by

UNIT-III chapter-2 Remaing Topics

t-Distribution (OR) student's t-distribution - Small Samples

Introduction :- when the size of the sample (n) less them 30, then that sample is called a small sample, ie $n<30$, the sample is called a small sample. Sampling distribution approches a normal distribution and values of simple statistic are considered best estimates of the parameters in a population in case of large samples we can use normal distribution to test for a specified population mean (or) equality of means, in small samples only if the sample is drawn from a normal population whose standard deviation is known it involves the concept of "degrees of freedom".

→ The different types of distributions that involve sized samples where ($n<30$) are

i) t-distribution ii) f-distribution.

Degree's freedom :- The number of degrees of freedom is equal to the total number of observations less than the number of independent constraints imposed on the observations.

It is denoted by ϑ.

Def:- t-distribution :-

If $x_1, x_2, x_3 \cdots x_n$ be any random sample of size n drawn from a normal population with mean μ and variance σ^2, then the test statistic t is defined by

$t = \dfrac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$ where $\bar{x}$ = Sample mean

$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is an unbiased estimate of σ^2.

The test statistic $\boxed{t = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}$ is a random variable having the t-distribution with $\boxed{\nu = n-1}$ degrees of freedom and with probability density function

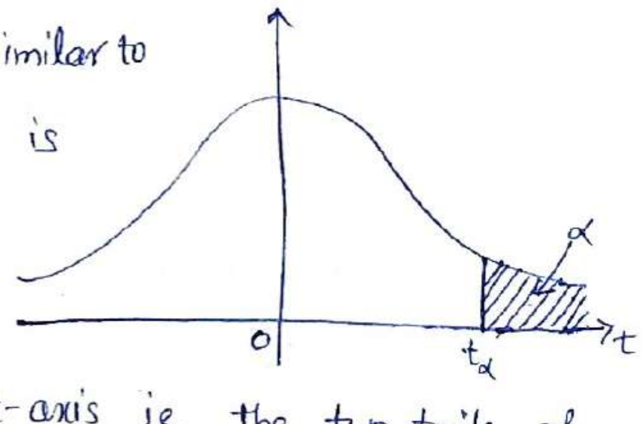
$f(t) = y_0\left(1+\dfrac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}$ where $\nu = n-1$ and y_0 is constant got by $\int_{-\infty}^{\infty} f(t)\,dt = 1$. This is known as "Student's t-distribution (or) Simply t-distribution".

<u>Note</u>:- t-distribution is used for testing of hypothesis when the sample size is small and population S.D σ is not known.

imp

Properties of t-distribution:

1) The Shape of t-distribution is ~~also~~ bell-shaped curve, which is similar to the normal distribution and is symmetrical about the mean.



2) The t-distribution curve is also asymptotic to the t-axis ie the two tails of the curve on both sides of $t=0$ extends to infinity.

Cont.

3) It is symmetrical about the line $t=0$.

4) The form of the probability curve varies with degrees of freedom ie with sample size.

5) It is Unimodal with mean = Median = Mode.

6) The mean of standard normal distribution and as well as t-distribution is zero but the variance of t-distribution depends upon the parameter ν which is called the degrees of freedom.

<u>Note</u> :- The selected values of t_α for various values of ν can be obtained from the table of t-distribution. where t_α denotes the area under t-distribution to its right is equal to α. These values are tabulated values.

→ In tables left-hand column contains value of ν, the column heading are areas α in the right-hand tail of the t-distribution, and the entries are values of t_α.

→ It is not necessary to tabulate values of t_α for $\alpha < 0.50$, as it follows from the symmetry of the t-distribution that $t_{1-\alpha} = -t_\alpha$ ie the t-value leaving area of $1-\alpha$ to the right and therefore an area α to its left, is equal to the negative t-value which leads an area α in the right.

Applications of the t-distribution

The t-distribution has a wide number of applications in statistics, some of them are given below

(i) To test the significance of the sample mean, when population variance is not given.

(ii) To test significance of the mean of the sample, ie to test if the sample mean differs significantly from the population mean.

(iii) To test the significance of the difference between two samples means (or) to compare two samples.

Problems :-

(1) Find (a) $t_{0.05}$ when $\nu = 16$

(b) $t_{-0.01}$ when $\nu = 10$

(c) $t_{0.995}$ when $\nu = 7$.

Sol:- from t-distribution table.

(a) when $\nu = 16$, $t_{0.05} = 1.746$

(b) when $\nu = 10$, $t_{-0.01} = -2.764$.

(c) when $\nu = 7$, $t_{0.995} = t_{1-0.995}$ $(\because t_{1-\alpha} = -t_{\alpha})$

$$= -t_{0.005}$$

$$\boxed{t_{0.995} = -3.449}$$

Uses of t- test

(1) To test for a specified mean.

(2) To test for equality of two means of two independent samples drawn from two normal populations, S.D of the populations being unknown.

(3) To test the significance of difference between the means of paired data.

Test of Hypothesis for Single Mean :-

Let a random sample size $n(<30)$ has a sample mean $\bar{x}$. To test the hypothesis that the population mean μ has specified value μ_0 when population S.D σ is not known.

Null Hypothesis $H_0 : \mu = \mu_0$

Alternative Hypothesis $H_1 : \mu \neq \mu_0$. (Two-tailed test).

Test statistic $\boxed{t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}}$ where s = Sample S.D

follows t-distribution with $\boxed{\nu = n-1}$ degrees of freedom.

Conclusion :- we Calucculate the value of $|t|$ and compare with table value of t at α level of significance.

→ If $|t| > t_\alpha$, then H_0 is Rejected.

→ If $|t| < t_\alpha$, then H_0 is Accepted.

→ Confidence interval (or) Fudicial Interval for one mean (μ) (Small Samples).

$$\boxed{\left(\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} , \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \right).}$$

(i) 95% confidence limits for μ is given by

$$\left(\bar{x} - t_{0.05} \cdot \frac{S}{\sqrt{n}} , \bar{x} + t_{0.05} \cdot \frac{S}{n} \right) \text{ with } \nu=(n-1) \text{ d.f}$$

(ii) 99% confidence limits (or) interval for μ is given by $\left(\bar{x} - t_{0.01} \frac{S}{\sqrt{n}} , \bar{x} + t_{0.01} \frac{S}{\sqrt{n}} \right)$.

imp
Note:- For a two-tailed test at α- level of significance, value of $\frac{\alpha}{2}$ is taken for α.

Problems :-

(1) The average breaking strength of the steel rods is to be specified to be 18.5 thousand pounds. To Test this sample of 14 rods were tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experiment significant?

Sol:- Given sample size $n = 14$.

Sample mean $\bar{x} = 17.85$

Sample S.D $s = 1.955$

population mean $\mu = 18.5$

Degrees of freedom $\nu = n-1 =$

Scanned with ACE Scanner

Null Hypothesis $H_0: \mu = 18.5$ (The result of the experiment is not significant)

Alternative Hypothesis $H_1: \mu \neq 18.5$

Level of significance $\alpha: 0.05$

Test statistic $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{17.85 - 18.5}{\frac{1.955}{\sqrt{13}}}$

$$t = \frac{0.65}{0.542} = -1.199.$$

$$|t| = |-1.199| = 1.199.$$

Tabulated value 't' at 5% level of significance for 13 d.f for two-tailed test $t_\alpha = 2.16$.

$\therefore |t| < t_\alpha$.

Then H_0 is Accepted.

(2) A random sample of six steel beams has a mean compressive strength of 58,392 p.s.i (pounds per square inch) with S.D of 648 p.s.i. Use this information and the level of significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i. Assume normality.

Sol: Given $n = 6$ $(n < 30)$ which is small sample.

$$\bar{x} = 58,392 \text{ p.s.i}$$

$$S = 648 \text{ p.s.i}$$

Degrees of freedom (d.f) $= n-1 = 6-1 = 5$.

In this problem σ is unknown and $n < 30$. Hence we use t-distribution.

Null Hypothesis H_0 : $\mu = 58,000$

Alternative Hypothesis H_1 : $\mu \neq 58,000$

Level of significance α : 0.05

Test statistic $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}}$

$$t = \frac{58,392 - 58,000}{\frac{648}{\sqrt{5}}} = 1.353$$

$$|t| = |1.353| = 1.353.$$

tabulated value at 0.05 level of two-tailed test of $\nu = 5$ d.f, $t_\alpha = 2.571$

$$\because \quad |t| < t_\alpha$$

Hence H_0 is Accepted.

ie The average compressive strength of the steel beam is not equal to 58,000 p.s.i.

H.W

1) A sample of 100 iron bars is said to be drawn from a large number of bars whose lengths are normally distributed with mean 4 feet and S.D 6ft. If the sample mean is 4.2 feet, can the sample be regarded as a truly random sample.

A sample of size 10 was taken from a population S.D of sample is 0.03. Find the maximum error with 99% confidence

Sol:- Given standard deviation $S = 0.03$

$n = 10$

maximum error $E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$

$t_{\frac{\alpha}{2}}$ for 99% of $\nu = 9$ d.f $= 3.25$

$\therefore E = 3.25 \times \frac{0.03}{\sqrt{10}}$

$E = 0.0325.$

4). A sample of 11 rats from a central population had an average blood viscosity of 3.92 with a S.D of 0.61. Estimate the 95% confidence limits for the mean blood viscosity of the population.

Sol:- Given $n = 11$

degrees of freedom $\nu = n-1 = 11-1 = 10.$

Confidence interval for mean μ is $\left(\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$

$\therefore \bar{x} = 3.92$

S.D is $S = 0.61$

$t_{\frac{\alpha}{2}}$ for 95% and $\nu = 11-1 = 10$ d.f $= 2.23$

$= \left(3.92 - 2.23 \times \frac{0.61}{\sqrt{11}}, 3.92 + 2.23 \frac{0.61}{\sqrt{11}}\right)$

$= (3.92 - 0.41, 3.92 + 0.41)$

$= (3.51, 4.33)$

(5) A sample of 15 members has a mean 67 and S.D 5.2. Is this sample has been taken from a large population of mean 70?

Sol: Given $n = 15$

$\bar{x} = 67$

$S = 5.2$

$\mu = 70$

Null Hypothesis $H_0: \mu = 70$

Alternative Hypothesis $H_1: \mu \neq 70$

Level of significance $\alpha: 0.05$ (Assume)

Test statistic $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}} = \dfrac{67 - 70}{\frac{5.2}{\sqrt{15-1}}}$

$t = \dfrac{-3}{\frac{5.2}{\sqrt{14}}} = \dfrac{-3}{\frac{5.2}{3.741}} = \dfrac{-3}{1.39}$

$t = 2.158$

$|t| = |2.158| = 2.158$

Tabulated value at 0.05 level of two-tailed test

of $\nu = 14$ d.f $t_\alpha = t_{\frac{0.05}{2}} = t_{0.025} = 2.145$

$|t| = 2.158$, $t_\alpha = 2.145$

$\therefore$ $|t| > t_\alpha$.

$\therefore$ H is Rejected

problems related to t-distribution (when Standard deviation of the sample is not given directly)

imp

1) The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data

Item :	1	2	3	4	5	6	7	8	9	10
Life in 1000 hrs	1.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

can we accept the hypothesis that the average life time of ~~bulbs~~ bulbs is 4000 hrs

Sol:- Given $n = 10$, $\mu = 4000$

Here S.D and sample mean is not given directly

We have to determine these S.D and mean as follows

Mean $\bar{x} = \frac{1.2 + 4.6 + 3.9 + 4.1 + 5.2 + 3.8 + 3.9 + 4.3 + 4.4 + 5.6}{10}$

$\bar{x} = \frac{41}{10} = 4.1 \approx 4100$

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

x	$x-\bar{x}$	$(x-\bar{x})^2$
1.2	-2.9	8.41
4.6	0.5	0.25
3.9	-0.2	0.04
4.1	0	0
5.2	1.1	1.21
3.8	-0.3	0.09
3.9	-0.2	0.04
4.3	0.2	0.04
4.4	0.3	0.09
5.6	1.5	2.25

$\Sigma(x-\bar{x})^2 = 12.42$

Scanned with ACE Scanner

$$\therefore S^2 = \sum_{i=1}^{10} \frac{(x_i-\bar{x})^2}{n-1} = \sum_{i=1}^{10} \frac{(x_i-\bar{x})^2}{10-1}$$

$$= \frac{1}{9}(12.42 = 1.38$$

$$S^2 = 1.38 \cong 1380$$

$$\cancel{S=}\sqrt{\cancel{1380}} \cancel{=1.174}.\ S = \sqrt{1380} = 37.148$$

Null Hypothesis $H_0: \mu = 4000$

Alternative Hypothesis $H_1: \mu \neq 4000$.

Level of Significance α : 0.05

Test statistic $t = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \dfrac{4100-4000}{\frac{\sqrt{1380}}{\sqrt{10-1}}}$

$$t = \frac{100}{\frac{\sqrt{1380}}{\sqrt{9}}} = \frac{100}{\frac{\sqrt{1380}}{3}} = \frac{100}{\frac{\sqrt{1380}}{3}}$$

$$\cancel{t=}\ \cancel{\frac{100}{37.058}} \quad t = \frac{100}{\frac{\sqrt{1380}}{3}}$$

$$t = \frac{100}{\frac{37.148}{3}} = \frac{100}{12.38} = 8.076.$$

$$|t| = |8.076| = 8.076.$$

$$\boxed{|t| = 8.076.}$$

tabulated value of t_α at 5% level of two-tailed test of $\vartheta = 10-1 = 9$ d.f

$$t_\alpha = 2.262$$

$\therefore$ $|t| > t_\alpha$

$\therefore$ H_0 is Rejected.

Scanned with ACE Scanner

A random sample of 10 boys had the following IQs: (7)

70, 120, 110, 101, 88, 83, 95, 98, 107 and 100

(a) Do these data support the assumption of a population mean IQ of 100!

(b) Find reasonable range in which most of the mean IQ values of samples of 10 boys lie.

Sol:- Given $n = 10$. , $\mu = 100$.

$$\bar{x} = \frac{70+120+110+101+88+83+95+98+107+100}{10}$$

$$\bar{x} = \frac{972}{10} = 97.2$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	−27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	−9.2	84.64
83	−14.2	201.64
95	−2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
		$\boxed{\Sigma(x-\bar{x})^2 = 1883.60}$

We have $s^2 = \sum_{i=1}^{10} \frac{(x_i-\bar{x})^2}{n-1} = \frac{1883.60}{10-1} = \frac{1883.60}{9}$

$$s^2 = 203.73$$

$$s = \sqrt{203.73} = 14.27.$$

Null Hypothesis H_0: $\mu = 100$

Alternative Hypothesis H_1: $\mu \neq 100$

Level of significance α: 0.05

Test statistic $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \dfrac{97.2 - 100}{\frac{14.27}{\sqrt{10-1}}}$

$$t = \frac{97.2-100}{\frac{14.27}{\sqrt{9}}} = \frac{-2.8}{\frac{14.27}{3}} = \frac{-2.8}{4.7566}$$

$$t = -0.5466$$

$$|t| = |-0.5466| = 0.5466.$$

Tabulated value at 5% level of significance of two-tailed test of $\nu = 10-1 = 9$ d.f $= 2.26$.

$\therefore$ $|t| < t_\alpha$.

$\therefore$ H_0 is Accepted.

(b) The 95% confidence limits $= \left(\bar{x} - t_{0.05}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.05}\frac{s}{\sqrt{n}}\right)$

$$= \left(97.2 - 2.26 \times \frac{14.27}{\sqrt{10}}, 97.2 + 2.26 \times \frac{14.27}{\sqrt{10}}\right)$$

$$= (97.2 - 2.26 \times 4.512, 97.2 + 2.26 \times 4.512)$$

$$= (97.2 - 10.198, 97.2 + 10.198)$$

$$= (107.4, 87).$$

Confidence limits = 107.4, 87.

The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% significance level assuming that for 9 degrees of freedom ($t = 1.833$ at $\alpha = 0.05$).

Sol:- Calculation for sample mean and S.D.

$$\text{Mean } \bar{x} = \frac{70+67+62+68+61+68+70+64+64+66}{10}$$

$$\bar{x} = \frac{660}{10} = 66.$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	4	16
67	1	1
62	-4	16
68	2	4
61	-5	25
68	2	4
70	4	16
[illegible]	8	4
64	-2	4
64	-2	0
66	0	

$$\Sigma(x-\bar{x})^2 = 90$$

$$s^2 = \sum_{i=1}^{10} \frac{(x_i-\bar{x})^2}{n-1} = \frac{90}{10-1} = \frac{\cancel{90}^{10}}{\cancel{9}} = 10$$

$$s = \sqrt{10} = 3.16.$$

Null Hypothesis $H_0: \mu = 64$

(The average height is not greater than 64 inches)

Alternative hypothesis $H_1: \mu > 64$. (Right-Tailed test)

Level of significance α: 0.05

Test statistic $t = \frac{\bar{x}-\mu}{S/\sqrt{n-1}} = \frac{66-64}{\frac{3.16}{\sqrt{10-1}}} = \frac{2}{\frac{3.16}{3}}$

$$t = \frac{2}{1.0533} = 1.898$$

$$|t| = |1.898| = 1.898.$$

Tabulated value t_α for $\nu = 10-1 = 9$ d.f at 5% level of significance of Right-Tailed test $t_\alpha = 1.833$.

$\therefore |t| > t_\alpha$.

$\therefore H_0$ is Rejected.

H.W.

(4) Eight students were given a test in STATISTICS and after one month coaching they were given another test of the similar nature. The following table gives the increase I their marks in the second test over the first test.

Student No.	1	2	3	4	5	6	7	8
Increase of Marks.	4	-2	6	-8	12	5	-7	2

Do the marks indicate that the students have gained from the coaching.

H.W

5) A random sample of 8 envelopes is taken from the letter box of a post office and their weights in grams are found to be: 12.1, 11.9, 12.4, 12.3, 11.5, 11.6, 12.1, and 12.4.

(i) Does this sample indicate at 1% level that the average weight of envelopes received at their post office is 12.35 gms.

(ii) Find 95% confidence limits for the mean weight of the envelopes received at that post office.

Scanned with ACE Scanner

STUDENT's t-Test for Difference of Means (9)

(OR)

t-Test for Two means

Let $\bar{x}$ and $\bar{y}$ be the means of two independent samples of sizes n_1 and n_2 ($n_1 < 30$, $n_2 < 30$) drawn from two normal populations having means μ_1 and μ_2.

To Test whether the two population means are equal.

Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative hypothesis H_1: $\mu_1 \neq \mu_2$

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ with $\boxed{\nu = (n_1 + n_2 - 2)}$ d.f.

where $S^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

imp Note:- 1) when standard deviation and mean not given directly

$$\bar{x} = \frac{\Sigma x_i}{n}, \quad \bar{y} = \frac{\Sigma y_i}{n}.$$

$$\boxed{S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}{n_1 + n_2 - 2}}$$

2) Confidence interval for two population means are

$$\left[(\bar{x} - \bar{y}) - t_\alpha \cdot S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\bar{x} - \bar{y}) + t_\alpha \cdot S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

Ex i) 99% confidence interval for two population means is

$$\left[(\bar{x} - \bar{y}) - t_{0.005} S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\bar{x} - \bar{y}) + t_{0.005} S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

(ii) 95% Confidence interval for two population means

$$\left((\bar{x}-\bar{y}) - t_{0.025}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ (\bar{x}-\bar{y}) + t_{0.025}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

with $\nu = (n_1+n_2-2)$ d.f.

<u>Problems</u> :- 1) Samples of two types of electric bulbs were tested for length of life and following data were obtained

Type-I	Type-II
Sample number $n_1 = 8$	$n_2 = 7$
Sample mean $\bar{x} = 1234$ hrs	$\bar{y} = 1036$ hrs
Sample S.D $s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warrant that type-I is superior to type-II regarding length of life.

<u>Sol</u>:- Since sample sizes are small and σ_1, σ_2 are not Known, we use t-test.

Let μ_1 and μ_2 be the two population means.

given $n_1 = 8$, $n_2 = 7$

$\bar{x} = 1234$, $\bar{y} = 1036$

$s_1 = 36$, $s_2 = 40$

Null Hypothesis H_0: The two types I and II of electric bulbs are identical.

ie $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$.

Level of significance α : 0.05.

Since two sample means $\bar{x}$ and $\bar{y}$ are given and also sample S.D S_1 and S_2 are given we use

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

where $S^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$

$S^2 = \dfrac{8(36)^2 + 7(40)^2}{8+7-2} = 1659.08$

$\therefore t = \dfrac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.39$

$|t| = |9.39| = 9.39$

degrees of freedom $d.f = n_1 + n_2 - 2 = 8 + 7 - 2 = 13$.

Tabulated value at 5% level of two tailed test

$t_\alpha = t_{0.05} = t_{\frac{0.05}{2}} = t_{0.025}$ at $\nu = 13$ d.f $= 2.160$.

$\therefore |t| > t_\alpha$.

$\therefore$ H_0 is Rejected.

we conclude that the two types I and II of electric bulbs are not identical.

imp

2). Two ~~horses~~ horses A and B were tested according to the time (in seconds) to run a perticular track with the following results

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whethere the two horses have the same running capacity.

Sol:- Given $n_1 = 7$, $n_2 = 6$.

we first compute the sample means and standard deviation

$$\bar{x} = \frac{28 + 30 + 32 + 33 + 33 + 29 + 34}{7} = \frac{219}{7} = 31.286$$

$$\bar{y} = \frac{29 + 30 + 30 + 24 + 27 + 29}{6} = \frac{169}{6} = 28.16.$$

$$S.D \quad S = \sqrt{\frac{1}{n_1+n_2-2} \sum_{i=1}^{n} (x_i-\bar{x})^2 + (y_i-\bar{y})^2}$$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	+.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
		$\sum(x-\bar{x})^2 = 31.4358$			$\sum(y-\bar{y})^2 = 26.8336$

$$\therefore S^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2 + (y_i-\bar{y})^2}{n_1+n_2-2}$$

$$= \frac{31.4358 + 26.8336}{7+6-2} = \frac{58.2694}{11} = 5.23$$

$$S = \sqrt{5.23} = 2.3.$$

Null Hypothesis $H_0: \mu_1 = \mu_2$ (11)

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ (Two-tailed test)

level of significance α: 0.05 (Assume)

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$t = \frac{31.286 - 28.16}{2.3\sqrt{\frac{1}{7} + \frac{1}{6}}} = 2.443$$

$|t| = |2.443| = 2.443.$

tabulated value of t for $\nu = 7+6-2 = 11$ d.f at 5% level of significance of two-tailed test $t_\alpha = 2.2$

$\therefore |t| > t_\alpha.$

$\therefore$ Ho is Rejected.

3). The IQ s (intellegence quotients) of 16 students from one area of a city showed a mean of 107 with a S.D of 10, while the IQ's of 14 students from another area of the city showed a mean of 112 with a S.D of 8. Is there a significant difference between the IQ's of the two groups at a 0.05 level of significance?

Sol:- Let μ_1 and μ_2 be the means of two populations.

Let the Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ (Two.-tailed test)

given $n_1 = 16$, $n_2 = 14$

$\bar{x}_x = 107$, $\bar{y} = 112$

$s_1 = 10$, $s_2 = 8$

$\therefore$ Level of significance $\alpha : 0.05$

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{16(10)^2 + (14)(8)^2}{16+14-2}$$

$S^2 = 89.14$

$S = \sqrt{89.14} = 9.44$

$\therefore t = \dfrac{107 - 112}{9.44\sqrt{\frac{1}{16} + \frac{1}{14}}} = -1.447$

$|t| = |-1.447| = 1.447.$

tabulated value of t for $\nu = n_1 + n_2 - 2 = 16+14-2 = 28$ d.f

at 5% level of significance of two-tailed test

$t_\alpha = 1.701$

$\therefore |t| < t_\alpha.$

$\therefore$ Ho is Accepted.

Ex 17) Two independent samples of 8 and 7 items respectively had the following values. (12)

Sample I	11	11	13	11	15	9	12	14
Sample II	9	11	10	13	9	8	10	-

Is the difference between the means of samples significant?

Sol:- Given $n_1 = 8$, $n_2 = 7$

$$\bar{x} = \frac{11+11+13+11+15+9+12+14}{8} = \frac{96}{8} = 12$$

$$\bar{y} = \frac{9+11+10+13+9+8+10}{7} = \frac{70}{7} = 10.$$

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
11	-1	1	9	-1	1
11	-1	1	11	1	1
13	1	1	10	0	0
11	-1	1	13	3	9
15	3	9	9	-1	1
9	-3	9	8	-2	4
12	0	0	10	0	0
14	2	4			

$\boxed{\Sigma(x-\bar{x})^2 = 26}$ $\boxed{\Sigma(y-\bar{y})^2 = 16}$

$$S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma (x_i-\bar{x})^2 + (y_i-\bar{y})^2\right]$$

$$= \frac{1}{8+7-2}\left[26+16\right] = \frac{42}{13} = 3.23$$

$\therefore \quad S = 1.8$

Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ (Two-tailed-Test)

level of significance α: 0.05 (Assume)

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{12-10}{3.23\sqrt{\frac{1}{8}+\frac{1}{7}}}$

$$t = 2.15$$

$$|t| = |2.15| = 2.15.$$

tabulated value of t for $\nu = 8+7-2 = 13$ d.f at 5% level of significance of two-tailed test $t_\alpha = 2.16$.

$\therefore \quad |t| < t_\alpha$.

$\therefore$ H_0 is Accepted.

H.W

(5) To examine the hypothesis that the husbands are more intelligent than the wives, an investigator work a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows.

Husbands	117	105	97	105	123	109	86	78	103	107
wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with a ~~spe~~ reasonable test at the level of significance of 0.05.

Scanned with ACE Scanner

Paired - Sample t-Test

If $(x_1, y_1), (x_2, y_2), (x_3, y_3) \cdots (x_n, y_n)$ be the pairs of sales data before and after the sales promotion in a business concern, we apply paired t-test to examine the significance of the difference of the two situations.

Let $d_i = x_i - y_i$ (or) $y_i - x_i$ for $i = 1, 2, 3 \cdots$

Let Null Hypothesis $H_0 : \mu_1 = \mu_2$ (ie $\mu = 0$) (There is no difference between the means in two situations).

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Test statistic $\boxed{t = \dfrac{\bar{d} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{\bar{d}}{s/\sqrt{n}}}$ $(\mu = 0)$.

where $\bar{d} = \sum_{i=1}^{n} \frac{d_i}{n}$, $s^2 = \sum_{i=1}^{n} \frac{(d_i - \bar{d})^2}{n-1}$

above statistic follows $\boxed{\nu = n-1}$ d.f

problems :-

(1) Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below

Before :	67	24	57	55	63	54	56	68	33	[illegible]
After :	70	38	58	58	56	67	68	75	42	[illegible]

Test whether the intensive training is useful at 0.05 level of significance.

Sol:- from the given ~~data~~ data, we see that we are to use paired t-test

let μ be the mean of population

let the Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

Level of significance α: 1% (or) 0.01

test statistic $t = \frac{\bar{d}}{s/\sqrt{n}}$.

Caluculation for $\bar{d}$ and S

B.P Before intake of drug (x)	B.p after intake of drug(y)	$d_i = y - x$	$(d_i - \bar{d})$	$(d_i - \bar{d})^2$
110	120	10	8	64
120	118	-2	-4	16
123	125	2	0	0
132	136	4	2	4
125	121	-4	-6	36
		$\Sigma d_i = 10$		$\Sigma(d_i - \bar{d})^2 = 120$

$$\bar{d} = \sum_{i=1}^{n} \frac{d_i}{n} = \bar{d} = \frac{\cancel{10}^{2}}{\cancel{5}} = 2$$

$$\boxed{\bar{d} = 2}$$

$$S^2 = \sum_{i=1}^{n} \frac{(d_i - \bar{d})^2}{n-1} = \frac{\cancel{120}^{30}}{\cancel{4}} = 30$$

$$S = \sqrt{30} = 5.477$$

$$t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{2}{\frac{5.477}{\sqrt{5}}} = 0.82$$

$$|t| = |0.82| = 0.82$$

tabulated value of t at 1% level of significance for two-tailed test of $\nu = 5-1 = 4$ d.f.

$$t_\alpha = t_{0.01} = t_{\frac{0.01}{2}} = t_{0.005}, \quad t_{0.005} \text{ of } \nu = 4 \text{ d.f}$$

from t-distribution table $t_\alpha = 4.604$

$\therefore \ |t| = 0.82, \quad t_\alpha = 4.604$

$\therefore \ |t| < t_\alpha$

$\therefore \ H_0$ is Accepted.

We conclude that there is no significant change in Blood pressure after intake of a certain drug.

H.W (3). Memory capacity of 10 students were tested before and after training. State whether the training was effective (or) not from the following scores.

Before training :	12	14	11	8	7	10	3	0	5	6
After training :	15	16	10	7	5	12	10	2	3	8

H.W (4) The average losses of workers, before and after certain ~~drug~~ program are given below. Use 0.05 level of significance to test whether the program is effective (Paired sample t-test). 40 and 35, 70 and 65, 45 and 42, 120 and 116, 35 and ~~65~~ 33, ~~and~~ 55 and 50, 77 and 73.

F- distribution

→ F-test is used to Test the equality of population variances

→ It is also used to test whether two independent samples have been drawn from the normal population with same variances.

Let S_1^2 and S_2^2 are two sample variances of samples of sizes n_1 and n_2 respectively drawn from the normal population, with variances σ_1^2 and σ_2^2. To determine whether the two samples come from two populations having equal variances.

Consider the sampling distribution of the ratio of the variances of the two independent samples defined by

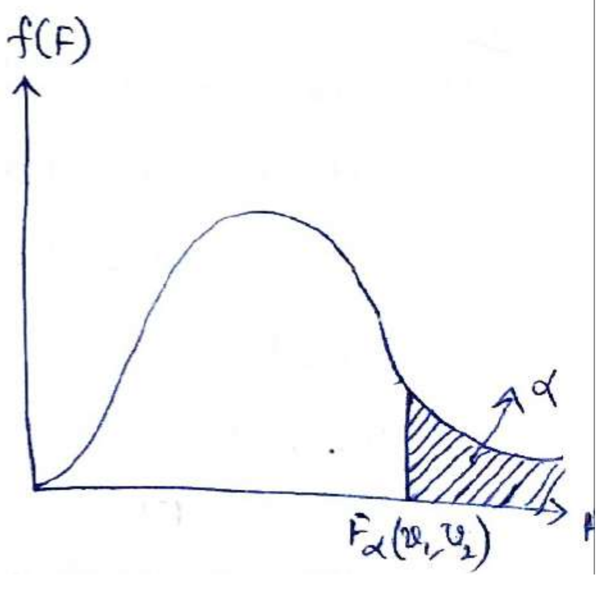
$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$ which follows

F-distribution with $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$ d.f.

Properties of F-distribution:

1) F - distribution curve lies entirely in first-quadrant.

2). F-distribution is free from population parameters and depends upon degrees of freedom only.



The F-curve depends not only on the two parameters ν_1 and ν_2 but also on the order in which they are stated.

4) The mode of F-distribution is less than unity.

5) $$F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_\alpha(\nu_2, \nu_1)}$$

where $F_\alpha(\nu_1, \nu_2)$ is the value of F with ν_1 and ν_2 degrees of freedom such that the area under the F-distribution curve to the right of F_α is α.

Example :- For an F-distribution, find

(a) $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$

(b) $F_{0.01}$ with $\nu_1 = 24$ and $\nu_2 = 19$

(c) $F_{0.95}$ with $\nu_1 = 19$ and $\nu_2 = 24$.

(d) $F_{0.99}$ with $\nu_1 = 28$ and $\nu_2 = 12$.

Sol :- From table F-distribution, (a) $F_{0.05}$ with $\nu_1 = 7$, $\nu_2 = 5$

$$F_{0.05} = 2.71.$$

(b) $F_{0.01}$ with $\nu_1 = 24$, $\nu_2 = 19$, $F_{0.01} = 2.92$

(c) $$F_{0.95}(19, 24) = \frac{1}{F_{0.05}(24, 19)} = \frac{1}{2.11} = 0.473$$

(d) $$F_{0.99}(28, 12) = \frac{1}{F_{0.01}(12, 28)} = \frac{1}{2.90} = 0.3448$$

Test for Equality of two population variances

Let two independent random samples of sizes n_1 and n_2 be drawn from two normal populations.

To test the hypothesis that the two population variances σ_1^2 and σ_2^2 are equal.

Null Hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

Alternative Hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$

Test statistic $F = \frac{S_1^2}{S_2^2}$ (or) $\frac{S_2^2}{S_1^2}$ according as $S_1^2 > S_2^2$ (or) $S_2^2 > S_1^2$ follows F-distribution with (n_1-1, n_2-1) d.f.

where $S_1^2 = \sum_{i=1}^{n_1} \frac{(x_i-\bar{x})^2}{n_1-1}$

$S_2^2 = \sum_{i=1}^{n_2} \frac{(y_i-\bar{y})^2}{n_2-1}$. where S_1^2 and S_2^2 are the variances of the two samples.

Conclusion :- If $|F| > F_\alpha$, H_0 is Rejected.

If $|F| < F_\alpha$, H_0 is Accepted.

Note :- 1) In numerical problems, we take the greater of the two variances S_1^2, S_2^2 in the numerator and the other in the denominator.

ie $F = \frac{\text{Greater variance}}{\text{Smaller variance}}$

Note ②:- $F_\alpha(\nu_1, \nu_2)$ is the value of F with ν_1 and ν_2 degrees of freedom such that under the F-distribution to the right of F_α is α. In tables F_α tabulated for $\alpha = 0.05$ and $\alpha = 0.01$ various combinatoions of the degrees of freedom ν_1 and ν_2. clearly value of F at 5% Significance is lower than at 1%.

Problems:-

(1) In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same variance.

Sol:- Let σ_1^2 and σ_2^2 be the variances of the two normal populations from which the samples are drawn.

Let the Null Hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$

given $n_1 = 8$, $n_2 = 10$.

$$\Sigma(x_i - \bar{x})^2 = 84.4, \quad \Sigma(y_i - \bar{y})^2 = 102.6$$

Level of significance α: 5%

$$\therefore S_1^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n_1 - 1} = \frac{84.4}{8-1} = \frac{84.4}{7} = 12.057$$

$$S_2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n_2 - 1} = \frac{102.6}{10-1} = \frac{102.6}{9} = 11.4$$

Here $S_1^2 > S_2^2$.

Then test statistic $F = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4}$

$$F = 1.057$$

$$|F| = |1.057| = 1.057.$$

Degrees of freedom (d.f) $\Rightarrow \nu_1 = n_1 - 1 = 8 - 1 = 7$

$\nu_2 = n_2 - 1 = 10 - 1 = 9$

Tabulated value of F at 5% level for (7, 9) d.f

$$F_\alpha = 3.29.$$

$$F_{0.05}(7, 9) = 3.29$$

$\therefore$ $|F| < F_\alpha$.

$\therefore$ Null Hypothesis H_0 is Accepted.

$\therefore$ we conclude that the populations have the same variance.

2) The time taken by workers in performing a job by method I and method II is given below

Method I	20	16	26	27	23	22	—
Method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly.

Let the Null Hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

Alternative Hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$

$\alpha: 0.05$ (Assume)

Here S_1^2 and S_2^2 are not given directly, so that

calculation of sample variances.

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
20	-2.3	5.29	27	-7.4	54.76
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16
			38	3.6	12.96.
		$\boxed{\Sigma(x-\bar{x})^2 = 81.34}$			$\boxed{\Sigma(y-\bar{y})^2 = 133.72}$

Given $n_1 = 6$, $n_2 = 7$

$$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{134}{6} = 22.3$$

$$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{241}{7} = 34.4$$

and $\Sigma(x-\bar{x})^2 = 81.34$, $\Sigma(y-\bar{y})^2 = 133.72$

$$S_1^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n_1-1} = \frac{81.34}{6-1} = \frac{81.34}{5} = 16.26$$

$$S_2^2 = \sum_{i=1}^{n} \frac{(y_i-\bar{y})^2}{n_2-1} = \frac{133.72}{7-1} = \frac{133.72}{6} = 22.29.$$

Since $S_2^2 > S_1^2$

then the statistic $F = \frac{S_2^2}{S_1^2} = \frac{22.29}{16.268} = 1.37$

Tabulated value of F for (n_1-1, n_2-1) d.f $(5,6)$

ie at 5% level of significance $F_\alpha = 4.39$.

$\therefore \; |F| < F_\alpha$.

$\therefore$ H_0 is Accepted.

$\therefore$ There is no significant difference between the variances of the distribution by the workers.

imp 3). The nicotine contents in milligrams in two samples of tobacco were found to be as follows.

Sample A	24	27	26	21	25	—
Sample B	27	30	28	31	22	36

can it be said that the two samples have come from the same normal population?

Sol:- To test whether the two samples came from same normal population, we have to test

(i) The equality of variances by using F-test.

(ii) The equality of means by using t- test.

Given $n_1 = 5$, $n_2 = 6$.

3> Caluculation for means and S.D's of the samples.

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
24	0.6	0.36	27	-2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	-1	1
21	3.6	12.96	31	2	4
25	0.4	0.16	22	-7	49
			36	7	49
$\boxed{\Sigma x_i = 123}$		$\boxed{\Sigma(x-\bar{x})^2 = 21.2}$	$\boxed{\Sigma y_i = 174}$		$\boxed{\Sigma(y-\bar{y})^2 = 108}$

$$\bar{x} = \frac{\Sigma x_i}{n_1} = \frac{123}{5} = 24.6$$

$$\bar{y} = \frac{\Sigma y_i}{n_2} = \frac{174}{6} = 29.$$

$$S_1^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n_1-1} = \frac{21.2}{4} = 5.3$$

$$S_2^2 = \sum_{i=1}^{n} \frac{\cancel{x_i-\bar{x}}(y_i-\bar{y})^2}{n_2-1} = \frac{108}{5} = 21.6$$

(i) <u>F - Test</u>

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$

Level of Significance α: 0.05 (Assume)

Here $S_2^2 > S_1^2$

Test statistic $F = \dfrac{S_2^2}{S_1^2} = \dfrac{21.6}{5.3} = 4.075$

$\therefore |F| = |4.075| = 4.075$

$\because$ Tabulated value of F for $(n_2-1, n_1-1) = (5,4)$ d.f at 5% level of significance of two-tailed test is

$$F_\alpha = 6.26$$

$\therefore \quad |F| < F_\alpha.$

$\therefore$ H_0 is Accepted.

$\therefore$ The variances are equal.

(ii) <u>t-test</u>

$\therefore$ Null Hypothesis H_0: $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$ (Two-tailed test)

Level of significance α: 0.05 (Assume)

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Given $\bar{x} = 24.6$, $\bar{y} = 29$

$\Sigma(x_i - \bar{x})^2 = 21.2$, $\Sigma(y_i - \bar{y})^2 = 108$

$$S^2 = \sum_{i=1}^{n} \frac{\left((x_i - \bar{x})^2 + (y_i - \bar{y})^2\right)}{n_1 + n_2 - 2}$$

$$S^2 = \frac{21.2 + 108}{5 + 6 - 2} = 14.35$$

Scanned with ACE Scanner

$\because S^2 = 14.35 \Rightarrow S = \sqrt{14.35} = 3.78$

$\therefore \boxed{S = 3.78}$

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$t = \frac{24.6 - 29}{3.78\sqrt{\frac{1}{5} + \frac{1}{6}}}$$

$$t = -1.92$$

$\therefore |t| = |-1.92| = 1.92$

tabulated value of t for $\nu = n_1 + n_2 - 2$

$\nu = 5 + 6 - 2 = 9$ d.f

at 5% level of significance of two-tailed test

$\therefore t_\alpha = t_{0.05} = t_{\frac{0.05}{2}} = t_{0.025}$ at 9 d.f

$t_\alpha = 2.26$

$\therefore |t| < t_\alpha$

$\therefore H_0$ is Accepted.

$\therefore$ The two means are equal.

$\therefore$ From (i) and (ii) we conclude that the two samples came from same normal population.

H.W
4). Two random samples reveal the following results:

Samples	Size	Sample mean	Sum of squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

H.W
5). An instructor has two classes A and B, in a particular subject. class A has 16 students while class B has 25 students. On the same examination, although there was no significant difference in mean grades, class A has standard deviation of 9 while class B had a standard deviation of 12. Can conclude ~~that~~ at the 0.01 level of significance that the variableility of class B is greater than that of A?

H.W
6). The measurements of the output two units have given the following results. Assuming that both samples have been obtained from the normal populations at 10% significant level, Test whether the two populations have the same variance.