MODULE - 4

Sampling distribution:

* **Population:**

A group of individual / elements / objects / items etc that we want to study is called population.

A population with finite no. of elements is called finite population.

A population having an infinite no. of elements is called infinite population.

* **Sample:**

If subset selected from the population to analyze and make inferences about the population is called sample.

* **Parameter:**

The statistical constants / measures (like mean (μ), variance (σ^2) etc) computed from the population are called parameters.

* **Statistic:**

The statistical constants are measures (like mean $(\bar{x})$, variance (s^2), ... etc) computed from the sample are called statistics.

Note: for population:

If the population data is $y_1, y_2, y_3, \ldots, y_N$

then mean is $\mu = \dfrac{\sum_{i=1}^{N} y_i}{N}$

variance is $(\sigma^2) = \dfrac{\sum (y_i - \mu)^2}{N}$

Standard deviation $(\sigma) = \sqrt{\dfrac{\sum (y_i - \mu)^2}{N}}$

these are parameters

Module-4

For sample:

Sample data is $x_1, x_2, x_3, \ldots, x_n$ then mean is

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

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

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

statistics

sampling distribution:

consider a population of size N. If we select a sample of sample size 'n' from the population, then the total no. of possible samples is given by

$K = N^n$ (with replacement)

(or):- $^{N}C_n = \frac{N!}{(N-n)! \cdot n!}$ (without replacement)

For each of these 'K' samples, we can calculate some statistics (t), in parameter mean($\bar{x}$), variance(s^2), standard deviation(s)... etc as shown in following table.

sample	statistic	If $t = \bar{x}$	If $t = s^2$
1	t_1	$\bar{x}_1$	s_1^2
2	t_2	$\bar{x}_2$	s_2^2
3	t_3	$\bar{x}_3$	s_3^2
⋮	⋮	⋮	⋮
k	t_k	$\bar{x}_k$	s_k^2

The set of values of statistic (t) obtained one or each sample constitute what is called sampling distribution of statistic (t).

Here statistic(t) is considered as a random variable taking the values $t_1, t_2, \ldots, t_k$ so far this distribⁿ of 't' also we can evaluate statistical measures in particular mean, variance, S.D...

Mean of sampling distribution of statistic (t):

$$E(x) \text{ or } \bar{t} = \frac{\sum_{i=1}^{k} t_i}{k}$$

variance of sampling distribution of statistic (t) is

$$var(t) = \frac{\sum_{i=1}^{k} (t_i - \bar{t})^2}{k}.$$

S.D of sampling distribution of statistic (t) is

$$SD(t) = \sqrt{var(t)} = \sqrt{\frac{\sum_{i=1}^{k} (t_i - \bar{t})^2}{k}}$$

The S.D (t) is called as standard error.

Note:

If $t = \bar{x}$, then mean of sampling distribution of means is $E(\bar{x}) = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$

variance of sampling distribution of means is

$$var(\bar{x}) = \frac{\sum_{i=1}^{k} (\bar{x}_i - E(\bar{x}))^2}{k}$$

S.D of sampling distribution of means.

$$S.D(\bar{x}) = \underbrace{\sqrt{var(\bar{x})}}_{\text{this is called standard error}} = \sqrt{\frac{\sum_{i=1}^{k} (\bar{x}_i - E(x))^2}{k}}$$

Eg: A population consists of 5 numbers these are 6, 8, 10, 12, 14. If all samples of size 2 are drawn from this population with replacement find

i) the total no. of possible samples ii) mean of population iii) variance of population iv) S.D of the population. v) All possible sample of size '2' vi) Sampling distribution of means vii) mean of sampling distribution of means.

viii) variance of sampling distribution of means

ix) standard error.

Sol: Given,

population is 6, 8, 10, 12, 14

population size is N = 5

sample size (n) = 2.

i) The total no. of possible samples of size 2 from the population with replacement is given by

$K = N^n = 5^2 = 25$.

ii) population mean is $\mu = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{6+8+10+12+14}{5}$

$= 10$.

iii) population variance is $\sigma^2 = \frac{\sum_{i=1}^{N} (y_i - \mu)^2}{N}$

$= [(6-10)^2 + (8-10)^2 + (10-10)^2 + (12-10)^2 + (14-10)^2] / 5$

$= [(-4)^2 + (-2)^2 + (2)^2 + (4)^2] / 5$

$= 16 + 4 + 4 + 16 / 5$

$= 10/5 = 2$.

iv) SD of the population.

$\sigma = \sqrt{8} = 2\sqrt{2}$.

v) All possible samples are

		2nd draw				
		6	8	10	12	14
1st draw	6	(6,6)	(6,8)	(6,10)	(6,12)	(6,14)
	8	(8,6)	(8,8)	(8,10)	(8,12)	(8,14)
	10	(10,6)	(10,8)	(10,10)	(10,12)	(10,14)
	12	(12,6)	(12,8)	(12,10)	(12,12)	(12,14)
	14	(14,6)	(14,6)	(14,10)	(14,12)	(14,14)

Total = 25.

vi) The sampling distribution of means ($\bar{x}$)

All possible samples of size '2' are

(6,6)(6,8)(6,10)(6,12)(6,14)
(8,6)(8,8)(8,10)(8,12)(8,14)
(10,6)(10,8)(10,10)(10,12)(10,14)
(12,6)(12,8)(12,10)(12,12)(12,14)
(14,6)(14,8)(14,10)(14,12)(14,14)

6	7	8	9	10
7	8	9	10	11
8	9	10	11	12
9	10	11	12	13
10	11	12	13	14

Mean of sampling distribution of means is

$$E(\bar{x}) = \frac{\Sigma \bar{x}_i}{25}$$

$$E(\bar{x}) = \frac{6+7+8+9+10+7+8+9+10+11+8+9+10+11+12+9+10+11+12+13+10+11+12+13+14}{25} = \frac{250}{25}$$

$= 10$

vii) variance of sampling distribution of means is

$$var(\bar{x}) = \frac{\Sigma(\bar{x}_i - E(\bar{x}))^2}{25}$$

$$= \frac{\begin{aligned}&(6-10)^2+(7-10)^2+(8-10)^2+(9-10)^2+(10-10)^2\\&+(7-10)^2+(8-10)^2+(9-10)^2+(10-10)^2+(11-10)^2\\&+(8-10)^2+(9-10)^2+(10-10)^2+(11-10)^2+(12-10)^2\\&+(9-10)^2+(10-10)^2+(11-10)^2+(12-10)^2+(13-10)^2\\&+(10-10)^2+(11-10)^2+(12-10)^2+(13-10)^2+(14-10)^2\end{aligned}}{25}$$

$= (6-10)^2 + 2(7-10)^2 + 3(8-10)^2 + 4(9-10)^2 + 4(11-10)^2 + 3(12-10)^2 + 2(13-10)^2 + (14-10)^2 / 25$

$= (-4)^2 + 2(-3)^2 + 3(-2)^2 + 4(-1)^2 + 4(1)^2 + 3(2)^2 + 2(3)^2 + (4)^2 / 25$

$= 2(4)^2 + 4(3)^2 + 6(2)^2 + 8(1)^2 / 25$

$= 2(16) + 4(9) + 6(4) + 8(1) / 25 = 32+36+24+8/25$

$= 100/25 = 4.$

viii) standard deviation of sampling distribution of mean is

$$S.D(\bar{x}) = \sqrt{Var(\bar{x})} = \sqrt{4} = 2.$$

ix) ∴ standard error = $S.D(\bar{x}) = 2$.

prob: A population consists of five ~~members~~ numbers 2, 3, 6, 8, 11. If all samples of 2 are drawn from this population without replacement. Then find

i) mean of population.
ii) variance of population
iii) S.D of population.
iv) all possible samples of size 2.
v) sampling distribution of mean.
vi) mean, variance, standard error of this sampling distribution of means.

Sol: Given,

population: 2, 3, 6, 8, 11

N = 5. (population size)
n = 2 (sample size)

Total no. of possible samples is $K = {}^{N}C_{n} = {}^{5}C_{2}$

$$= \frac{5!}{3!\,2!} = \frac{5\times4}{2} = 10.$$

These 10 samples are,

(2,3) (2,6) (2,8) (2,11)
(3,6) (3,8) (3,11)
(6,8) (6,11)
(8,11)

Sampling distribution of means $(\bar{x})$

2.5, 4, 5, 6.5

4.5, 5.5, 7

7, 8.5.

9.5.

mean of sampling distribution is

$$E(\bar{x}) = \frac{\Sigma \bar{x}_i}{10}$$

$= 2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5/10$

$= 60/10$

$= 6.$

variance of sampling distribution of mean

$$var(\bar{x}) = \frac{\Sigma [\bar{x}_i - E(\bar{x})]^2}{10}$$

$$= \frac{(2.5-6)^2 + (4-6)^2 + (5-6)^2 + (6.5-6)^2 + (4.5-6)^2 + (5.5-6)^2 + (7-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2}{10}$$

$= (3.5)^2 + (2)^2 + (1)^2 + (0.5)^2 + (1.5)^2 + (0.5)^2 + (1)^2 + (1)^2 + (1.5)^2 + (3.5)^2 / 10$

$= 2(3.5)^2 + (2)^2 + 3(1)^2 + 2(0.5)^2 + 2(1.5)^2 / 10$

$= 24.5 + 4 + 3 + 0.5 + 4.5 / 10$

$= 36.5/10 = 3.65$.

standard deviation of sampling distribution is

$S.D(\bar{x}) = \sqrt{var(\bar{x})} = \cancel{\sqrt{4.05}} = \sqrt{3.65}$

$= 1.9104$.

$\therefore$ Standard error $= SD(\bar{x}) = 1.9104$

Standard error of sampling distribution of some well known statistics:

case-1: When samples are drawn randomly from an infinite population.

1] standard error of sampling distribution of mean ($\bar{x}$) is $SE(\bar{x}) = \frac{\sigma}{\sqrt{x}}$ (or) $\sqrt{\sigma^2/n}$

2] Standard error of sampling distribution of proportion (p) is $SE(P) = \sqrt{\frac{PQ}{n}}$

3] SE of sampling distribution of S.D (S)

$SE(S) = \frac{\sigma}{\sqrt{2n}}$ (or) $\sqrt{\frac{\sigma^2}{2n}}$.

4] SE of sampling distribution of difference of means ($\bar{x}_1 - \bar{x}_2$) is

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{x_1} + \frac{\sigma_2^2}{x_2}}.$$

5] SE of sampling distribution of difference of proportions means ($P_1 - P_2$) is

$$SE(P_1 - P_2) = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$

6]. SE of sampling distribution of difference of standard deviation ($S_1 - S_2$) is

$$SE(S_1 - S_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}.$$

Case-1: When samples are drawn randomly from an infinite population.

1] $S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ (or) $\sqrt{\frac{\sigma^2}{n}}$

2] $S.E(P) = \sqrt{\frac{PQ}{n}}$

3]. $S.E(S) = \frac{\sigma}{\sqrt{2n}}$ (or) $\sqrt{\frac{\sigma^2}{2n}}$

4] $S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Here,

n = Sample size

σ = population S.D.

$\bar{x}$ = Sample mean

p = Sample proportion

P = population proportion.

$\therefore Q = 1-P$

Here $\bar{x}_1$ = Sample1 mean

$\bar{x}_2$ = Sample 2 mean

σ_1 = population 1 standard deviation

σ_2 = population 2 standard deviation

n_1 = sample size 1

n_2 = samplesize 2

5]. $S.E(P_1 - P_2) = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$

P_1 = Sample 1 proportion.

P_2 = Sample 2 proportion

P_1 = population 1 proportion

P_2 = population 2 proportion

3. $S.E(S_1-S_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$

Here S_1 = sample 1 standard deviation

S_2 = sample 2 standard deviation.

case-2: when samples are drawn randomly from a finite proportion without replacement

$\rightarrow S.E(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

$\rightarrow S.E(P) = \sqrt{\frac{PQ}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Estimation:

Estimation of population parameters by using corresponding sample statistics.

1. point estimation: A particular value or a single value of a statistic which is used to estimate the unknown value corresponding population parameter is called point estimation.

2. unbiased estimation: A sample statistic (t) is said to be unbiased estimate of corresponding population parameter (θ). if $E(\bar{x}) = \theta$

That is if the mean of sampling distribution of statistic 't' is equal to corresponding population parameter (θ)

Note:

1) The sample mean $\bar{x}$ is an unbiased estimate of population mean μ ($\because E(\bar{x}) = \mu$)

2) The sample proportion (P) is an unbiased estimate of population proportion P. ($\because E(P) = P$).

3]. The sample variance $(s^2) = \frac{\sum(x_i - \bar{x})^2}{n}$ is not an unbiased estimate of population variance σ^2

$(\therefore E(s^2) \neq \sigma^2)$.

But S^2 defined by $S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} \longrightarrow$ (2) is an unbiased estimate of σ^2

from (1) & (2), we get

$$ns^2 = (n-1)S^2$$

$$s^2 = \left(\frac{n-1}{n}\right)S^2$$

$$s^2 = \left[1 - \frac{1}{n}\right]S^2$$

$\therefore$ if n is large, $\frac{1}{n} \rightarrow 0$ then we get $s^2 = S^2$.

Thus for large n, s^2 is an unbiased estimate of σ^2.

central limit theorem:

If 't' is a any statistic, then the standardized statistic Z defined by $Z = \frac{t - E(t)}{S.E(t)}$ follows normal distribution with mean 0 and standard deviation 1 for sufficiently large n.

In particular: If $t = \bar{x}$, then

$Z = \frac{\bar{x} - E(\bar{x})}{S.E(\bar{x})} = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ follows normal distribution for sufficiently large n.

ii) If $t = p$ then

$Z = \frac{p - E(p)}{S.E(p)} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ follows normal distribution

with mean 0 & S.D 1.

(iii) If $t = s$, then

$$z = \frac{s - E(s)}{S.E(s)} = \frac{s - \sigma}{\left(\frac{\sigma}{\sqrt{2n}}\right)}$$ follows normal

distribution with mean '0' and S.D '1' for sufficiently large n.

Estimation:

1. point estimation
2. Interval estimation

In this case, we determine two constants C_1, C_2 s.t

$P(C_1 < \theta < C_2$ for a given value of statistic $t) = 1 - \alpha$

Here $[C_1, C_2]$ is called as $(1-\alpha)100\%$ confidence interval at α level of significance for estimating the unknown value of population parameter θ. By using corresponding sample.

statistic t

i) confidence Limits for estimating μ using $\bar{x}$:

by central limit theorem, we known that

$$z = \frac{\bar{x} - E(\bar{x})}{S.E(\bar{x})} = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$$

follows normal distribution with mean as '0' and S.D '1' for large values of n.

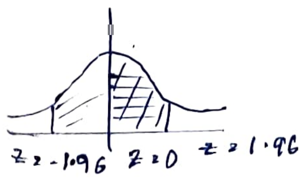
By using the area property of normal distribution

we have $P(-1.96 \le z \le 1.96) = 0.95$

$= 2P(0 \le z \le 1.96)$

$= 2[0.4750]$

$= 0.95$



$$\Rightarrow P\left(-1.96 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 1.96\right) = 0.95$$

$$\Rightarrow P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} \le \bar{x}-\mu \le 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\Rightarrow P\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\Rightarrow \cancel{\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}}$$

95% confidence Limit interval for estimating $\mu = \left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$

$\therefore$ 95% confidence limit for $\mu = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

Similarly 99% confidence Limit for $\mu = \bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$

99.73% ... $= \bar{x} \pm 3\left[\frac{\sigma}{\sqrt{n}}\right]$

90% ... $= \bar{x} \pm 1.64\left[\frac{\sigma}{\sqrt{n}}\right]$

In general $= \bar{x} \pm Z_{\frac{\alpha}{2}}\ SE(\bar{x})$

$Z_{\frac{\alpha}{2}}$ → critical value of z

2] confidence Limits for estimating population proportion P using sample proportion p.

95% confidence Limits for estimating P using p

$= p \pm \cancel{2.58}\ 1.96\ S.E(p)$

95% 99% $= p \pm 2.58\ SE(p)$

99.73% $= p \pm 3\ SE(p)$

90% $= p \pm 1.64\ SE(p)$

Here $S.E(p) = \sqrt{\frac{PQ}{n}}$

Confidence limits for $(\mu_1 - \mu_2)$:

95% ..... $= (\bar{x}_1 - \bar{x}_2) \pm 1.96 \; SE(\bar{x}_1 - \bar{x}_2)$

99% ..... $= (\bar{x}_1 - \bar{x}_2) \pm 2.58 \; SE(\bar{x}_1 - \bar{x}_2)$

99.73% ..... $= (\bar{x}_1 - \bar{x}_2) \pm 3 \; SE(\bar{x}_1 - \bar{x}_2)$

90% ..... $= (\bar{x}_1 - \bar{x}_2) \pm 1.64 \; SE(\bar{x}_1 - \bar{x}_2)$

where $SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\dfrac{\sigma_1^2}{x_1} + \dfrac{\sigma_2^2}{x_2}}$

Confidence lts for $P_1 - P_2$:

95% ..... $= (P_1 - P_2) \pm 1.96 \; SE(P_1 - P_2)$

99% ..... $= (P_1 - P_2) \pm 2.58 \; SE(P_1 - P_2)$

99.73% ..... $= (P_1 - P_2) \pm 3 \; SE(P_1 - P_2)$

90% ..... $= (P_1 - P_2) \pm 1.64 \; SE(P_1 - P_2)$.

where $SE(P_1 - P_2) = \sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}$.

Pb: A sample of size 100 is drawn from a large population. If sample mean is 7.4 kg & S.D is 1.2, then find 95% & 99% confidence limits for estimating population mean.

for 95% $\Rightarrow \mu = \bar{x} \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$

for 99% $\Rightarrow \mu = \bar{x} \pm 2.58 \dfrac{\sigma}{\sqrt{n}}$

Given $n = 100$, $\bar{x} = 7.4$, $\sigma = 1.2$.

95% $\Rightarrow \mu = 7.4 \pm 1.96 \left[\dfrac{1.2}{10}\right] = 7.4 \pm 1.96\,[0.12]$

$= 7.4 \pm 0.2352$

$= 7.1648 - 7.6352$

99% $\Rightarrow \mu = 7.4 \pm \overset{2.58}{\cancel{1.96}} \left[\dfrac{1.2}{10}\right] = 7.4 \pm 2.58\,[0.12]$

$= 7.4 \pm 0.3096 = 7.0904 / 7.7096$.

Sol: Given:

$n = 100$ (sample size)

$\bar{x} = 7.4$ (sample mean)

$s = 1.2$ (sample S.D)

Now we need to find 95% & 99% confidence limits for ~~ex~~ estimating population mean μ

95% confidence limits for μ

$= \bar{x} \pm 1.96\, S.E(\bar{x})$

$= \bar{x} \pm 1.96 \left[\frac{\sigma}{\sqrt{n}}\right]$

$= \bar{x} \pm 1.96 \left[\frac{s}{\sqrt{n}}\right]$ [Here s is unbiased estimate of σ for sufficiently large n]

$= 7.4 \pm 1.96 \left[\frac{1.2}{\sqrt{100}}\right]$

$= [7.1648, 7.6352]$

99% confidence limits for μ

$= \bar{x} \pm 2.58 \left[\frac{\sigma}{\sqrt{n}}\right]$ $\sigma \to s$

$= 7.4 \pm 2.58 \left[\frac{1.2}{\sqrt{100}}\right]$

$= [7.0904, 7.7096]$

Prob: A random sample of 700 units from a large consignment showed that 200 were damaged. Find i) 95% ii) 99% confidence limits for the proportion of damaged units in the consignment.

Sol: Given, $n = 700$ (sample size) ≥ 30 (large sample)

proportion of damaged items in the sample is

$(p) = \frac{200}{700} = \frac{2}{7}$.

$q = 1 - P = 1 - \frac{2}{7} = \frac{5}{7}$.

Now we find:

95% confidence limits for population proportion (P)

$= P \pm 1.96\ S.E(P)$

$= P \pm 1.96 \sqrt{\frac{PQ}{n}}$

$= p \pm 1.96 \sqrt{\frac{pq}{n}}$ ∵ Since p is unbiased estimate of P sly q is Q.

$= \frac{2}{7} \pm 1.96 \sqrt{\frac{10}{49} \times \frac{1}{700}}$

$= \frac{2}{7} \pm 1.96\ (0.01707)$ Ans: [0.2523, 0.3191]

$= 0.2857 \pm 0.334572$

$= [-0.0488,\ 0.6202]$

99% confidence limits.

$= p \pm 2.58 \sqrt{\frac{pq}{n}}$

$= \frac{2}{7} \pm 2.58 \sqrt{\frac{10}{49} \times \frac{1}{700}} =$

$= 0.2857 \pm 2.58\ (0.0170) = 0.2857 \pm 0.0441$

$= [0.24167,\ 0.3298]$

Prob: The mean height of students in university is 155cm and S.D is 15. what the probability that the mean height of sample of 36 students is less than 157 students.

Sol: Given, population mean $\mu = 155$

population S.D $\sigma = 15$.

sample size $n = 36\ (\geq 30)$ (large sample)

Now we find,

$P(\bar{x} < 157)$

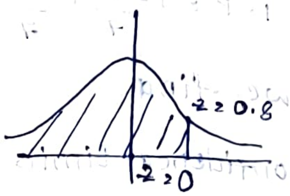
using,

$Z = \frac{t - E(t)}{SE(t)}$

$[$If $\bar{x} = 157 \Rightarrow \frac{157 - 155}{15/6} = \frac{2 \times 6}{15} = \frac{12}{15} = \frac{4}{5} = 0.8]$

$= \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{\bar{x} - 155}{(15/\sqrt{36})}$

Here $P(\bar{x} < 157) = P(z < 0.8)$



$= 0.5 + P(0 < z < 0.8)$

$= 0.5 + 0.2881$

$= 0.7881$

Prob: A normal population has a mean 0.1 & S.D 2.1. Find the probability that mean of a sample of size 900 will be negative.

Sol: x $\mu = 0.1$ (population mean)

$\sigma = 2.1$ (S.D)

$n = 900$ (sample size)

We need to find:

$P(\bar{x} < 0)$

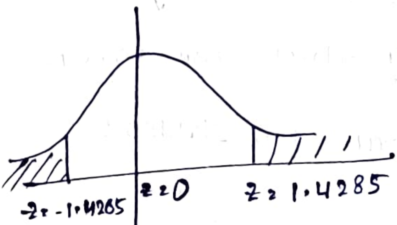
$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{0 - 0.1}{\frac{2.1}{30}} = \frac{-0.1 \times 30}{2.1} = \frac{-3}{2.1} = \frac{-30}{21}$$

$$= -\frac{10}{7}$$

$$= -1.4285$$

$P(\bar{x} < 0) = P(z < -1.4285)$

$= P(z > 1.4285)$



$= 0.5 - P(0 < z < 1.4285)$

$= 0.5 - 0.4236$

$= 0.0764$.

Testing of Hypothesis:

(Test of significance) [for Large Samples $(n \geq 30)$].

1] Define Null hypothesis (H_0)

A definite statement given about the population parameter is called null hypothesis.

It is usually a hypothesis of no difference (or)

Null hypothesis can be ~~at~~ in any of the following forms.

i) $\mu = \mu_0$, ~~if~~ ii) $P = P_0$ iii) $\mu_1 = \mu_2$ iv) $P_1 = P_2$

2] Define alternate hypothesis (H_1):

A hypothesis which is different from null hypothesis is called as alternate hypothesis.

If null hypothesis is $H_0 : \mu = \mu_0$, then the alternate hypothesis can be any of the following.

i) $H_1 : \mu \neq \mu_0 \rightarrow$ Two tailed alternate

In this case the test is known as two tailed test.

ii) $H_1 : \mu > \mu_0 \rightarrow$ right tailed alternate.

In this case the test is known as right tailed test.

iii) $H_1 : \mu < \mu_0 \longrightarrow$ Left tailed alternate

In this case the test is known as left tailed test.

3]. Fix a suitable level of significance (α).

then obtain critical value of x at this level of significance (α)

That is Single tailed test.

Find Z_α for right tailed / left tailed test.

Find $Z_{\alpha/2}$ for two tailed test.

4]. compute the value of test statistic.

$$Z = \frac{t - E(t)}{S.E(t)}.$$

5] i) Single tailed test.

compare the value of $|Z|$ and Z_α.

If $|Z| < Z_\alpha$, then we say that null hypothesis H_0 is accepted and alternate hypothesis H_1 is rejected.

2. If $|Z| > Z_\alpha$ then we say that H_0 is rejected and H_1 is accepted.

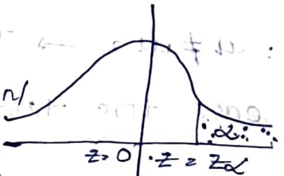
(ii) **Two tailed test:**

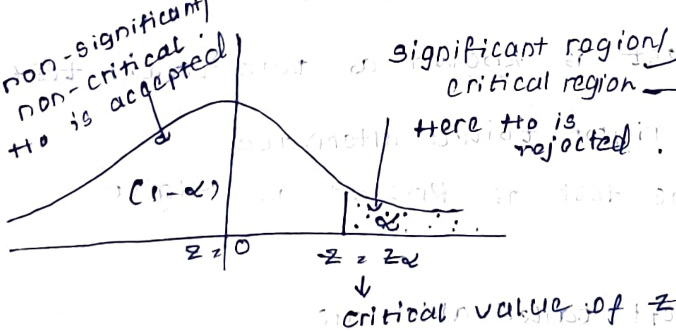
compare the value of $|Z|$ and Z_α.

1. If $|Z| < Z_{\alpha/2}$, then H_0 is accepted & H_1 is rejected.
2. If $|Z| > Z_{\alpha/2}$, then H_0 is rejected & H_1 is accepted.

Finding critical values of Z at level of significance α.

1] In right tailed test the critical value of Z denoted by Z_α at α level of significance is given by $P(Z > Z_\alpha) = \alpha \longrightarrow ①$.





Eg: If $\alpha = 1\% = 0.01$, then we get from I

$P(Z > Z_\alpha) = 0.01$

Sol: $P(Z > Z_\alpha) = 0.01$

$0.5 - P(0 < Z < Z_\alpha) = 0.01$

$P(0 < Z < Z_\alpha) = 0.49$.

$Z_\alpha = 2.33$

thus,

for $\alpha = 1\% = 0.01$, $Z_\alpha = 2.33$.

for $\alpha = 2\% = 0.02$, $Z_\alpha = 2.05$

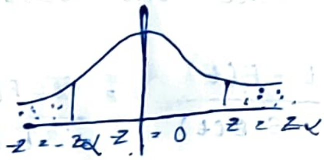
for $\alpha = 5\% = 0.05$, $Z_\alpha = 1.645$

for $\alpha = 10\% = 0.1$, $Z_\alpha = 1.28$.

2). In left tailed test the critical value of z denoted by z_α at α level of significance is given by $P(z < -z_\alpha) = \alpha$.

this is same as,

$$P(Z > z_\alpha) = \alpha$$



∴ The critical value of z are same.

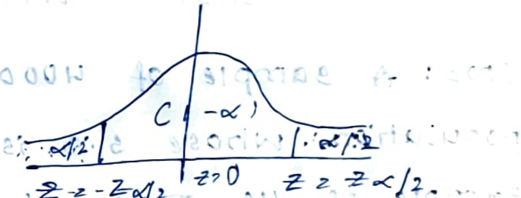
3). In two tailed test, the critical value of z, denoted by $z_{\frac{\alpha}{2}}$, at α level of significance is given by

$$P(|z| < z_{\alpha/2}) = \alpha$$

$$\Rightarrow P(z > z_{\alpha/2}) + P(z < -z_{\alpha/2}) = \alpha.$$

$$\Rightarrow 2P(z > z_{\alpha/2}) = \alpha$$

$$\Rightarrow \boxed{P(z > z_{\alpha/2}) = \frac{\alpha}{2}}$$



Eg: If $\alpha = 5\%$, then find $P(z > z_{\alpha/2}) = \frac{0.05}{2} = 0.025$

$$P(z > z_{\alpha/2}) = 0.025$$

$$0.5 - P(z > z_{\alpha/2}) = 0.025$$

$$P(z > z_{\alpha/2}) = 0.5 - 0.025 = 0.475$$

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

thus,

for $\alpha = 5\% = 0.05$, $z_{\alpha/2} = 1.96$.

$\alpha = 1\% = 0.01$. $z_{\alpha/2} = 2.58$

$\alpha = 2\% = 0.02$, $z_{\alpha/2} = 2.33$

$\alpha = 10\% = 0.1$, $z_{\alpha/2} = 1.645$

Critical values of z

	Level of significance			
	$\alpha = 1\%$	2%	5%	10%
Two tailed	2.58	2.33	1.96	1.645
Right	2.33	2.05	1.645	1.28
left	2.33	2.05	1.645	1.28

case - I : test of significance for single mean.

In this case, the test statistic is

$$Z = \frac{t - E(t)}{S.E(t)} = \frac{\bar{x} - E(t)}{S.E(t)} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

This can be used

i) To check the significant difference b/w $\bar{x}$ & μ.

ii) To check whether $\mu = \mu_0$

iii) To check whether the given sample is drawn from the population with mean μ has a specified value μ_0 (that is with $\mu = \mu_0$)

Prob: A sample of 4000 items is taken from a population whose S.D is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38 at 5%. level of significance.

Sol: Given

Sample size is $n = 400$

population S.D is $\sigma = 10$

sample mean is $\bar{x} = 40$

population mean $\mu = 38$

level of significance = ?

$\alpha = 5\% = 0.05$.

1. Null hypothesis H_0:

The sample has come from the population with mean $\mu = 38$.

2. Alternate hypothesis H_1: $\mu \neq 38$.

(Two tailed test)

3. Level of significance: $\alpha = 5\% = 0.05$.

the corresponding critical value of Z is $Z_{\alpha/2}$:

$Z_{\alpha/2} = 1.96$ (for 2-tailed test at $\alpha = 5\%$)

4. The value of test statistic is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{40 - 38}{10/20} = \frac{2}{1/2} = 4.$$

5. Here

$|Z| = |4| = 4$

$Z_{\alpha/2} = 1.96$

$|Z| > Z_{\alpha/2}$

Ho is rejected

That is: The sample has not come from the population with mean $\mu = 38$.

Prob: According to the norms established for a mechanical aptitude test, persons who are 18 years old have an avg height of 73.2 with S.D of 8.6. If 49 randomly selected persons of that age averaged 76.7, test the hypothesis $\mu = 73.2$ against the alternate hypothesis $\mu > 73.2$ at 1% level of significance.

Sol: Given, sample size is $n = 49$ (> 30) large sample

population S.D. is $\sigma = 8.6$

sample mean is $\bar{x} = 76.7$

population mean $\mu = 73.2$

1. Null hypothesis H_0: $\mu = 73.2$

2. Alternate hypothesis H_1: $\mu > 73.2$ (right tailed alternate)

3. Level of significance $\alpha = 1\% = 0.01$

the corresponding critical value of Z is

$Z_\alpha = 2.33$ (for $\alpha = 1\%$)

4. The value of test statistic is

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\cancel{40-38}\quad 76.7 - 73.2}{8.6/7}$$

$$= \frac{3.5 \times 7}{8.6}$$

$$= \frac{24.5}{8.6}$$

$$= 2.8488.$$

5. Here

$|Z| = |2.8488| = 2.8488.$

$Z_\alpha = 2.33$

$\Rightarrow$ Here $|Z| > Z_\alpha$

$\Rightarrow$ Null hypothesis $H_0 : \mu = 73.2$ is rejected.

Prob: In a large lot of electric bulbs, the mean life & S.D of bulbs are 350 hrs & 90 hrs respectively. A sample of 625 bulbs is choosen. It is found that the mean life & S.D of bulbs in the sample are 355 hrs & 90 hrs respectively. Can we conclude that the sample is drawn from given population at 5% level of significance.

population mean = μ = 360

population S.D = σ = 90

sample size n = 625.

sample mean = $\bar{x}$ = 355

sample S.D = 90.

$\alpha = 5\%$.

1. $H_0 = 360.$

2. H_1 : $\mu \neq 360$ (2 tailed)

3. $\alpha = 5\%$

$Z_{\alpha/2} = 1.96$

4. $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{355 - 360}{90/\sqrt{625}} = \dfrac{-5}{90/25} = \dfrac{-125}{90} = -1.3888...$

$= -1.389$

5. Here

$|Z| = |-1.389| = 1.389$

$Z_{\alpha/2} = 1.96$

→ Here $|Z| < Z_{\alpha/2}$

→ H_1 is rejected.

Case-2: Test of significance for single proportion.

In this case, the test statistic

$$Z = \frac{t - E(t)}{S.E(t)} = \frac{p - E(p)}{S.E(p)} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

This can be used.

i) To check the significant difference b/w p and P

p → sample proportion

P → population proportion

ii) To check whether population proportion P has a specified value P_0 That is to check whether $P = P_0$.

Prob: Experience had shown that 20% of a manufactured product is of the top quality. In 1 day's production of 400 articles only 50 are of top quality. Test the hypothesis at 5% level of significance.

Sol: Given sample size $n = 400$ (≥ 30) large sample.

Proportion of top quality articles in the sample

$$p = \frac{50}{400} = \frac{1}{8}$$

Population proportion $P = 20\% = \frac{20}{100} = \frac{1}{5} = 0.2$.

(Proportion of top quality articles in the population)

then $Q = 1 - P = 1 - \frac{1}{5} = \frac{4}{5}$

1. Null hypothesis H_0: 20% of manufactured product is of top quality. i.e $P = \frac{1}{5}$

2. Alternate hypothesis H_1: $P \neq \frac{1}{5}$ (two tailed alternate)

3. Level of significance $\alpha = 5\% = 0.05$

At 5% level of significance for 2 tailed test the critical value of z is $z_{\alpha/2} = 1.96$.

4. value of test statistic.

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{1}{8} - \frac{1}{5}}{\sqrt{\frac{4}{25} \times \frac{1}{400}}} = \frac{\frac{5-8}{40}}{\frac{2}{5 \times 20}}$$

$$= \frac{-3}{\cancel{40}_2} \times \frac{5 \times \cancel{20}}{2}$$

$$= \frac{-15}{4} = -3.75.$$

5. Here $|Z| = |-3.75| = 3.75$.

$Z_{\alpha/2} = 1.9$.

$\rightarrow |Z| > Z_{\alpha/2}$

$\rightarrow$ Null hypothesis H_0 is rejected.

i.e 20% of manufactured product is not of top quality.

prob: A manufacturer claimed that atleast 95% of the equipment which is supplied to a factory confirmed to specifications. An examination of 200 pieces of equipment revealed that 18 were faulty. Test this claim at 5% level of significance.

Sol: Given sample size $n = 200$ (>30) Large sample.

proption of equipment, confirm to specifications

$p = \frac{182}{200} = \frac{91}{100}$ $\rightarrow$ ~~population~~ sample proportion

population proportion of equipment confirmed to specification $P = 95\% = \frac{95}{100}$

$Q = 1 - P = 1 - \frac{95}{100} = \frac{5}{100}$

1. Null hypothesis H_0: 95% of equipment confirmed to specifications i.e $P = \frac{95}{100}$

2. Alternation hypothesis H_1: $P < \frac{95}{100}$ (left ~~right~~ tailed alternate)

3. $\alpha = 5\%$, $Z_\alpha = 1.645$

4. $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{91}{100} - \frac{95}{100}}{\sqrt{\frac{95}{100} \times \frac{5}{100} \times \frac{1}{200}}} = \frac{-\frac{4}{100}}{\sqrt{\frac{475}{2 \times 10^6}}}$

$= \frac{-4}{100} \times \frac{10^3}{15.411}$

$= -40 / 15.411 = -2.5955$

$= -2.6$

Y. Sai Samanvita
2024029840

5. Here $|Z| = |-2.61| = 2.6$

$Z_\alpha = 1.645$.

$\rightarrow |Z| > Z_\alpha$.

$\rightarrow$ H_0 is rejected.

i.e ~~Atleast~~ 95% confirmed to specifications is not accepted.

Prob: In a big city 325 men out of 600 were found to be smokers. Does this information support the claim that the no. of smokers are greater than the no. of non-smokers in the city.

SOL: Given,

sample size $n = 600$ (≥ 30) Large sample

Sample proportion $p = \frac{325}{600} = \frac{65}{120} = \frac{13}{24} = 0.5416$

population proportion,

$P = 50\% = \frac{50}{100} = \frac{1}{2}$

then $Q = 50\% = \frac{1}{2}$.

1. Null hypothesis H_0 : $P = \frac{1}{2}$

2. Alternate hypothes H_1 : $P > 50\%$ (right tailed)

$P > \frac{1}{2}$.

3. Level of significance 5%. $\alpha = 0.05$

$Z_\alpha = 1.645$.

4. test statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5416 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{600}}}$

$= \frac{0.0416}{\sqrt{\frac{1}{2400}}} = \frac{0.0416}{0.000416}$

$= 100$

$\rightarrow |Z| > Z_\alpha$. H_0 rejected.

5. Here $|Z| = |-2.61| = 2.6$

$Z_\alpha = 1.645$.

$\rightarrow |Z| > Z_\alpha$.

$\rightarrow$ H_0 is rejected.

i.e Atleast 95% confirmed to specifications is not accepted.

Prob: In a big city 325 men out of 600 were found to be smokers. Does this information support the claim that the no. of smokers are greater than the no. of non-smokers in the city.

Sol: Given,

sample size $n = 600$ (≥ 30) Large sample.

Sample proportion $p = \frac{325}{600} = \frac{65}{120} = \frac{13}{24} = 0.5416$

population proportion,

$P = 50\% = \frac{50}{100} = \frac{1}{2}$.

then $Q = 50\% = \frac{1}{2}$.

1. Null hypothesis H_0 : $P = \frac{1}{2}$

2. Alternate hypothes H_1 : $P > 50\%$ (right tailed)

$P > \frac{1}{2}$.

. Level of significance 5%, $\alpha = 0.05$

$Z_\alpha = 1.645$.

4. test statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5416 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{600}}}$

$= \frac{0.0416}{\sqrt{\frac{1}{2400}}} = \frac{0.0416}{0.000416} = 100$

$\rightarrow |Z| > Z_\alpha$, H_0 rejected.

Case-3: Test of significance for difference of means.

In this case, the test statistic is

$$Z = \frac{t - E(t)}{S.E(t)} = \frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \longrightarrow (1)$$

Note: 1] If population means are same or equal that is if $\mu_1 = \mu_2$ then from (1) we get.

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

If σ_1, σ_2 are not known, then we use their unbiased estimates s_1, s_2 respectively then (2) becomes

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \longrightarrow (3)$$

2] [If population means are equal that is if $\mu_1 = \mu_2$ then from (1), we get

$$Z = \bar{x}_1 - \bar{x}_2] \times$$

If the given random samples have been drawn from the same population that is if

$\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$

then we get from (1),

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \longrightarrow (4)$$

Here if σ is not known, then we use its unbiased estimate given by (based on the 2 samples)

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

Prob: The sizes & and means of 2 independent random samples are 400, 225 ; 3.5, 3 respectively. can we conclude that the samples are drawn from the same population with S.D 1.5. test this claim at 1% level of significance

Sol: Given,

$n_1 = 400$ $\quad \bar{x}_1 = 3.5$ $\quad \sigma = 1.5$

$n_2 = 225$. $\quad \bar{x}_2 = 3$. $\quad \alpha = 1\% = 0.01$

1. Null hypothesis H_0: The samples are drawn from same popn $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$.

2. Alternate hypothesis H_1: $\mu_1 \neq \mu_2$ (2 tailed)

3. Level of significance: $\alpha = 1\%$

$Z_{\alpha/2} = 2.58$.

4. Test Statistic: $Z = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$

$$Z = \frac{3.5 - 3}{\sqrt{\frac{2.25}{400} + \frac{2.25}{225}}} = \frac{0.5}{\sqrt{\frac{225}{4 \times 10^4} + \frac{225}{225 \times 10^2}}} = \frac{0.5}{\sqrt{\left[\frac{15}{2 \times 10^2} + \frac{1}{10}\right]}}$$

$$= \frac{0.5}{\sqrt{0.005625 + 0.01}} = \frac{0.5}{\sqrt{0.015625}}$$

$$= \frac{0.5}{\sqrt{\frac{1}{64}}} = 4.$$

5. $|Z| = 4$ $\Rightarrow |Z| > Z_{\alpha/2}$, H_0 is rejected.

$Z_{\alpha/2} = 2.58$

That is given samples have not been drawn from the same population with S.D $\sigma = 1.5$.

Prob: The no. of students in a class is 100. The avg marks scored by 36 girls is 70 with a S.D of 8, while the avg marks scored by 64 boys is 66 with a S.D of 10. Test at 1% level of significance whether girls perform better than boys.

Sol: $n_1 = 36$ (girls) $\quad \bar{x}_1 = 70 \quad \sigma = 10 \quad s_1 = 8$

$n_2 = 64$ (boys) $\quad \bar{x}_2 = 66 \quad \alpha = 1\% \quad s_2 = 10$

1. Null hypothesis: $\bar{x}_1 = \bar{x}_2$, $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$ (Both girls & boys perform equally well)

2. Alternate hypothesis H_1: $\mu_1 > \mu_2$
$\bar{x}_1 > \bar{x}_2$ (girls performed better than boys)

3. Level of significance: $\alpha = 1\%$ $\quad$ Right tailed

$Z_\alpha = 2.33$

4. Test statistic, $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$Z = \frac{70-66}{\sqrt{\frac{64}{36} + \frac{100}{64}}} = \frac{4}{\sqrt{3.3402}} = \frac{4}{1.8276} = 2.1886$$

5. $|Z| = |2.1886| = 2.1886$

$Z_\alpha = 2.33$

$Z_\alpha > |Z|$

H_0 is accepted

That is given samples, girls perform better than boys.

the given claim is verified.

Case - 4: Test of significance for difference of proportions.

In this case, the test statistic is,

$$Z = \frac{t - E(t)}{S.E(t)} = \frac{(P_1 - P_2) - E(P_1 - P_2)}{S.E(P_1 - P_2)}$$

$$= \frac{(P_1 - P_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \longrightarrow (1)$$

If the population proportions are equal i.e. $P_1 = P_2 = P$

then we get from (1)

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}} \longrightarrow (2).$$

$$= \frac{P_1 - P_2}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

If P is unknown then we use its unbiased estimate (based on 2 samples) given by

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}. \qquad Q = 1 - P.$$

Prob: A random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level of significance.

Sol: Given,

$n_1 = 400$ (men). $P_1 = \frac{200}{400} = \frac{1}{2}$ (proportion of men in favour of the proposal in sample 1)

$n_2 = 600$ (women) $P_2 = \frac{325}{600} = \frac{65}{120} = \frac{13}{24}$

(proportion of women in favour of proposal in sample 2)

1) Null hypothesis (H_0): $P_1 = P_2 = P$ proportions of men & women in favour of proposal are same.

2) Alternate hypothesis (H_1): $P_1 \neq P_2$ (two tailed alternate)

3) Level of significance (α): $\alpha = 5\% = 0.05$

$Z_{\alpha/2} = 1.96$

4). Test statistic: $Z = \dfrac{P_1 - P_2}{\sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}}$

since P is unknown we use its unbiased estimate given by: $P = n_1P_1 + n_2P_2 / n_1 + n_2$

$$Z = \frac{\frac{1}{2} - \frac{13}{24}}{\sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}} = -0.4166$$

$$P = \frac{400\left[\frac{1}{2}\right] + 600\left[\frac{13}{24}\right]}{1000} = \frac{200 + 325}{1000} = 0.525$$

$Q = 0.475$

$$Z = \frac{-0.4166}{\sqrt{0.249375\left[\frac{1}{400} + \frac{1}{600}\right]}} = \frac{-0.4166}{\sqrt{0.249375 \times 0.004166}}$$

$$= \frac{-0.4166}{0.03223} = -1.29$$

5) Here $|z| = |-1.29| = 1.29$

$Z_{\alpha/2} = 1.96 \Rightarrow |z| < Z_{\alpha/2}$

that is: The population proportion in favour of proposal are same.

Prob: In a city A 20% of random sample of 900 school boys have a certain slight physical defect. In another city B 18.5% of a random sample of 1600 school boys have the same defect. Is the difference b/w proportions significant at 5% LOS.

Sol: Given.

$$\left.\begin{array}{l} n_1 = 900 \\ P_1 = \frac{20}{100} \end{array}\right] \text{sample 1}$$

$$\left.\begin{array}{l} n_2 = 1600 \\ P_1 = \frac{18.5}{100} \end{array}\right] \text{sample 2.}$$

1] Null hypothesis (H_0): $P_1 = P_2 = P$.
(No difference in proportions)

2] Alternate hypothesis (H_1): $P_1 \neq P_2$
(Two tailed)

3] Level of significance (α): $\alpha = 5\% = 0.05$.

$Z_{\alpha/2} = 1.96$

4) Test statistic: $Z = \dfrac{P_1 - P_2}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{900\left[\frac{20}{100}\right] + 1600\left[\frac{18.5}{100}\right]}{2500}$$

$$= \frac{180 + 296}{2500} = \frac{476}{2500} = 0.1904.$$

$Q = 1 - P = 1 - 0.1904 = 0.8096$.

$$z = \frac{\frac{20}{100} - \frac{18.5}{100}}{\sqrt{0.1904 \times 0.8096\left[\frac{1}{900}+\frac{1}{1600}\right]}} = \frac{\frac{1.5}{100}}{\sqrt{0.15414784[0.001736]}}$$

$$= \frac{1.5/100}{\sqrt{0.0002676}}$$

$$= \frac{1.5/100}{0.01635} = 0.9169.$$

5]. $|z| = |0.9169| = 0.9169.$

$Z_{\alpha/2} = 1.96.$

$\Rightarrow Z_{\alpha/2} > |Z|$ [Ho is accepted]

That is: there is no significant ~~The~~ difference b/w proportions.

Small Sample Test: $(n<30)$.

1. t-Test for single Mean.

Assumptions:

i) The parent population from which sample is drawn is normally distribution.

ii) population standard deviation (σ) is not known. In this case, the test statistic is $\boxed{t = \frac{\bar{x}-\mu}{S/\sqrt{n}}}$

$$z = \frac{t - E(t)}{SE(t)} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

It does not follow normal distribution it follows t distribution with $v = n-1$ degrees of freedom.

Here S^2 is given by $S^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2$.

(or)

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

$\sim t_{v=n-1}$

$$\rightarrow \frac{S}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}$$

Prob: The heights of 8 ~~names~~ males in an athletic championship are found to be 175cm, 168cm, 165cm, 170cm, 167cm, 160cm, 173cm and 168cm. Can we conclude that the avg height is greater than 165cm. Test at 5% level of significance.

Sol: Sample size = $n = 8$ (< 30). So, it is small sample

degrees of freedom $(v) = n-1 = 8-1 = 7$.

Sample mean is: $\bar{x} = \frac{\Sigma x_i}{n} = \frac{175+168+165+170+167+160+173+168}{8}$

$= 168.25$

Sample standard deviation $(s) = ?$

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

$s^2 = \frac{1}{8}[(6.75)^2 + (0.25)^2 + (3.25)^2 + (1.75)^2 + (1.25)^2 + (8.25)^2$
$(4.75)^2 + (0.25)^2]$

$= \frac{1}{8}[(6.75)^2 + 2(0.25)^2 + (3.25)^2 + (1.75)^2 + (1.25)^2 + (8.25)^2 + (4.75)^2]$

$s^2 = 18.9375 \Rightarrow s = 4.$

To check claim:

1. Null hypothesis (H_0): $\mu = 165$
2. Alternate hypothesis (H_1): $\mu > 165$ (right tailed test).
3. Level of significance $(\alpha) = 5\% = 0.05$

The corresponding critical value of t is $t_v(\alpha)$

$t_v(\alpha) = 1.895$

$(v = 7, \alpha = 5\%$, right tailed)

4. Test statistic: $Z = t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{168.25 - 165}{\sqrt{\frac{18.9375}{7}}}$

$Z = \frac{3.25}{\sqrt{2.75357}} = 1.976$

Here $|t| = 1.976$, $t_v(\alpha) = 1.895$.

$\Rightarrow |t| > t_v(\alpha)$.

$\Rightarrow$ H_0 is rejected, H_1 is accepted.

$\therefore \mu > 165$ is accepted.

Prob: A random sample of 17 values from a normal population has a mean of 105cm and the sum of squares of deviations from this mean is 1225 sq.cm. Is this assumption of a mean 110cm for a normal population reasonable. Test at 5% level of significance. also obtain 95% confidence limits.

Sol: Given, $n = 17$ $(< 30) \rightarrow$ small sample

$v = n - 1 = 16$

$\bar{x} = \frac{\sum x_i}{n} = 105\text{cm}$.

$\sum (x_i - \bar{x})^2 = 1225$

$s = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} = \sqrt{\frac{1}{17}(1225)} = \sqrt{72.0588}$
$= 8.4887$.

To check the claim:

1. Null Hypothesis (H_0): $\mu = 110$.
2. Alternate Hypothesis (H_1): $\mu \neq 110$. (2 tailed)
3. Level of significance $(\alpha) = 5\%$.

$t_v(\alpha/2) = 2.12$ (critical value of t)
at $\alpha = 5\%$, $v = 16$ d.f.

4. Test Statistic: $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{105 - 110}{8.4887/\sqrt{16}} = \frac{-5}{8.4887/4}$

$= \frac{-20}{8.4887} = -2.3560$.

$|t| = |-2.3560| = 2.3560$

$t_v(\alpha/2) = 2.12$.

$\Rightarrow |t| > t_v(\alpha/2)$

$\Rightarrow$ H_0 is rejected, H_1 is accepted.

$\Rightarrow$ $\mu = 110$ is not reasonable.

95% confidence limits for estimating unknown μ using known $\bar{x}$

$$\Rightarrow \bar{x} \pm t_v(\alpha/2)\left(\frac{s}{\sqrt{n-1}}\right)$$

$$\Rightarrow 105 \pm 2.12\left[\frac{8.4887}{\sqrt{17-1}}\right] = 105 \pm 2.12\left[\frac{8.4887}{4}\right]$$

$$\Rightarrow (100.501,\ 109.499)$$

Case-2: t-test for difference of means

Assumptions:

1] The parent populations from which the random sample are drawn are normally distributed.

2] Population variances σ_1^2, σ_2^2 are equal and unknown

i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (σ is unknown)

In this case the test statistic is

$$t = \frac{(\bar{x}-\bar{y}) - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

- If population means are equal ($\mu_1 = \mu_2$) then we get

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\sigma^2\left[\frac{1}{n_1} - \frac{1}{n_2}\right]}}$$

Since σ^2 is not known, then we get to use its unbiased estimate (based on 2 samples) given by

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

$$= \frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2}$$

The test statistic is $\boxed{t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}}$

- It follows t distribution with $v = n_1+n_2-2$ degrees of freedom.

Prob: The following two independent random samples of sizes 8 and 7 are given below.

Sample 1 : 11 11 13 11 15 9 12 14

Sample 2 : 9 11 10 13 9 8 10

Is the difference between the sample means significant test at 5% level of significance.

Sol: Given,

Sample 1 $\quad\quad$ Sample 2

$n_1 = 8 \quad\quad n_2 = 7$

degrees of freedom$(v) = n_1 + n_2 - 2$

$= 8 + 7 - 2$

$= 13.$

$$\bar{x} = \frac{\Sigma x_i}{n} = (11+11+13+11+15+9+12+14)/8$$

$= 96/8$

$= 12.$

$$\bar{y} = \frac{\Sigma y_i}{n} = (9+11+10+13+9+8+10)/7$$

$= 70/7$

$= 10.$

$$\Sigma(x_i - \bar{x})^2 = (1)^2 + (1)^2 + (1)^2 + (1)^2 + (3)^2 + (3)^2 + (0)^2 + (2)^2$$

$= 1+1+1+1+9+9+0+4$

$= 26$

$$\Sigma(y_i - \bar{y})^2 = (1)^2 + (1)^2 + (0)^2 + (3)^2 + (1)^2 + (2)^2 + (0)^2$$

$= 1+1+0+9+1+4+0$

$= 16.$

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

$$= \frac{26+16}{13}$$

$$= \frac{42}{13}$$

$$= 3.2307$$

$$s = \sqrt{3.2307} = 1.\cancel{7}974$$

To check the claim:

1. Null Hypothesis (H_0) : $\bar{x} = \bar{y}$
2. Alternate Hypothesis (H_1) : $\bar{x} \neq \bar{y}$ (Two tailed)
3. Level of significance : $5\% = \alpha$

$$t_{13}(\alpha/2) = 2.160$$

4. Critical value $t = \dfrac{\bar{x}-\bar{y}}{\sqrt{s^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$

$$= \frac{12-10}{\sqrt{(3.2307)\left[\frac{1}{8}+\frac{1}{7}\right]}}$$

$$= \frac{2}{\sqrt{(3.2307)(0.2678)}}$$

$$= \frac{2}{\sqrt{0.8651}} = \frac{2}{0.9301}$$

$$= 2.15$$

5. ~~$|t| = |2.18241|$~~.

$|t| = |2.151| = 2.15$.

~~$|t| > t_v(\alpha/2)$~~ $\Rightarrow |t| < t_v(\alpha/2)$

$\Rightarrow H_0$ is accepted, H_1 is rejected

There is no significant difference b/w means.

* $(1-\alpha)100\%$ confidence limits

$$= (\bar{x} - \bar{y}) \pm t_v(\alpha/2)\left[\sqrt{S^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}\right]$$

<u>Prob</u>: The IQ's 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 b/w the IQ's of 2 groups ? Test at 5% Level of significance.

Sol: Sample-1

$n_1 = 16$

$\bar{x}_1 = 107$

S_1 ~~σ~~ $= 10$

Sample-2

$n_2 = 14$

$\bar{x}_2 = 112$

$S_2 = 8$

degrees of freedom $(v) = n_1 + n_2 - 2$

$= 16 + 14 - 2$

$= 28$.

$$S^2 = \frac{16[10]^2 + 14[8]^2}{28}$$

$$= \frac{16[100] + 14[64]}{28}$$

$$= \frac{1600 + 896}{28}$$

$$= \frac{2496}{28}$$

$$= 89.1428$$

To check the claim:

1] Null Hypothesis (H_0): $\bar{x}_1 = \bar{x}_2$

2] Alternate Hypothesis (H_1): $\bar{x}_1 \neq \bar{x}_2$ (two tailed)

3] level of significance $\alpha = 5\% = 0.05$

$t_v (\alpha/2) = 2.16$

4]. Critical value: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$

$$= \frac{107 - 112}{\sqrt{(89.1428)\left[\frac{1}{16} + \frac{1}{14}\right]}}$$

$$= \frac{-5}{\sqrt{(89.1428)(0.1339)}}$$

$$= \frac{-5}{\sqrt{11.9362}}$$

$$= -5 / 3.4548$$

$$= -1.4472$$

5] $|t| = |-1.44721| = 1.4472$

$t_v(\alpha/2) = 2.16$

$\Rightarrow t_v(\alpha/2) > |t|$

$\Rightarrow$ H_0 is accepted, H_1 is rejected.

$\therefore$ There is no significant difference.

Small sample Tests ($n < 30$)

Case-3: Paired t-Test

Assumptions: 1] sample sizes n_1 and n_2 are equal.

That is: $n_1 = n_2 = n$ (say).

2] Sample observations $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are not independent, but they are dependent in pairs.

In this case, the test statistic is $t = \dfrac{\bar{d}}{s_d/\sqrt{n-1}}$

Here $d = d_i = x_i - y_i$, $i = 1, 2, \ldots, n$

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{\Sigma (x_i - y_i)}{n} = \bar{x} - \bar{y}.$$

and $s_d^2 = \dfrac{\Sigma (d_i - \bar{d})^2}{n}$

It follows t-distribution $v = n-1$ d.o.f

$$s_d = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n}}$$

Prob: ~~An~~ IQ test was conducted to 5 persons before and after they were trained. The results are given below:

candidates :	I	II	III	IV	V.	
IQ's before training :	110	120	123	132	125.	(x_i)
IQ's after training :	120	118	125	136	121.	(y_i)

Test whether there is any change in IQ's after the training programme ?

Test at 1% level of significance.

Sol: Given

$n_1 = n_2 = n = 5.$

Also sample observations are dependent.

So, we can apply paired t-Test.

$$\bar{d} = \frac{\Sigma d_i}{n} = \bar{x} - \bar{y}.$$

computation table

x_i	y_i	$d = d_i = x_i - y_i$	$(d_i - \bar{d})^2$
110	120	-10	64
120	118	2	16
123	125	-2	0
132	136	-4	4
125	121	4	36
		$\Sigma d_i = -10$	$\Sigma (d_i - \bar{d})^2 = 120$

$$\bar{d} = \frac{-10}{5} = -2.$$

$$S_d = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n}}$$

$$= \sqrt{\frac{(-8)^2 + (4)^2 + (0)^2 + (-2)^2 + (6)^2}{5}}$$

$$= \sqrt{\frac{64 + 16 + 4 + 36}{5}}$$

$$= \sqrt{\frac{120}{5}} = \sqrt{24}$$

$$= 4.8989.$$

To verify the claim:

1] Null Hypothesis (H_0): [There is no significance difference in IQ's $\mu_1 = \mu_2$ (or) $\bar{x} = \bar{y}$]

(or) $\bar{d} = 0$.

2] ~~Alternate~~ Hypothesis (H_1):

$\bar{d} \neq 0$ (H_1) - two tailed - There is no significant difference in IQ's.

3]. Level of significance is

$\alpha = 1\% = 0.01$

critical value of t is $t_v(\alpha/2)$.

degrees of freedom $= n-1 = 5-1 = 4$

$t_v(\alpha/2) = 4.604$

4] ~~critical value~~

Test statistic $t = \dfrac{\bar{d}}{S_d / \sqrt{n-1}}$

$$t = \frac{-2}{4.8989/2}$$

$$= \frac{-4}{4.8989}$$

$$= -0.8165.$$

5]. $|t| = |-0.816| = 0.816.$

$t(\alpha/2) > |t|.$

$\Rightarrow$ Ho is accepted.

$\Rightarrow$ There is no significance difference in I.Q's.

Prob: The marks of 10 students obtained in tests before and after attending some coaching classes are given below.

Before: 54 76 92 65 75 78 66 82 80 78.

After: 60 80 86 72 80 72 66 88 82 73.

check whether students have benefitted from coaching at 5% level of significance?

$\bar{d} = \frac{\Sigma di}{n} = -\frac{13}{10}$

$= -1.3$

sol: Given,

$n_1 = n_2 = 10 = n$

computation table.

x_i	y_i	$d = d_i = x_i - y_i$	$(d_i - \bar{d})^2$
54	60	-6	22.09
76	80	-4	7.29
92	86	6	53.29
65	72	-7	32.49
75	80	-5	13.69
78	72	6	53.29
66	66	0	1.69
82	88	-6	22.09
80	82	-2	0.49
78	73.	5	39.69

$\Sigma d_i = -13$

$\Sigma (d_i - \bar{d})^2 = 246.1.$

$$S_d = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n}} = \sqrt{\frac{246.1}{10}} = \sqrt{24.61}$$

To verify the claim:

1] H_0: $\bar{x} = \bar{y}$ (or) $\bar{d} = 0$

2] H_1: $\bar{x} < \bar{y}$ (or) $\bar{d} < 0 \longrightarrow$ Left tailed test.

3]. L.O.S = α = 5%.

$$t_v(\alpha) = t_9(5\%) = 1.833.$$

4]. Test statistic

$$t = \frac{\bar{d}}{S_d / \sqrt{n-1}} = \frac{-1.3}{\sqrt{\frac{24.61}{9}}} = \frac{-1.3}{\sqrt{2.7344}}$$

$$= \frac{-1.3}{1.6536}$$

$$= -0.786.$$

5]. $|t| = |-0.786| = 0.786$

$\cancel{t_v(\alpha/2)}$ $t_v(\alpha) = 1.833$.

$|t| < t_v(\alpha)$

$\Rightarrow$ H_0 is accepted.

$\Rightarrow$ There is no significance difference.

Case-4: F-test [Small sample tests]

This test can be used to check the Significant difference between population variances. That is to check whether $\sigma_1^2 = \sigma_2^2$.

Assumptions:

1]. The parent population from which random samples are drawn are normally distributed.

2] Population variances σ_1^2 and σ_2^2 are unknown.

In this case: The test statistic is given by

$$F = \frac{\text{Greater variances}}{\text{Smaller variances}}$$

To use this:

First find S_1^2 and S_2^2 given by.

$$S_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x})^2 = \frac{n_1 s_1^2}{(n_1 - 1)}$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma (y_i - \bar{y})^2 = \frac{n_2 s_2^2}{(n_2 - 1)}$$

Here S_1^2 and S_2^2 are unbiased estimates of σ_1^2, σ_2^2 respectively.

$\Rightarrow$ If $S_1^2 > S_2^2$ then $F = \dfrac{S_1^2}{S_2^2}$.

It follows F-distribution with $V_1 = n_1 - 1$, $V_2 = n_2 - 1$ degrees of freedom

$\Rightarrow$ If $s_2^2 > s_1^2$, then $F = \frac{s_2^2}{s_1^2}$

It follows F-distribution with $v_1 = n_2 - 1$ and $v_2 = n_1 - 1$ degrees of freedom.

<u>Prob</u>: Two random samples of sizes 9 & 7, gave the sum of squares of deviations from their respective means as 175 and 95 respectively. Can they be regarded as drawn from normal populations with same variance. Test at 5% level of significance.

Sol: Given,

Sample 1: $n_1 = 9$, $\Sigma(x_i - \bar{x})^2 = 175$

Sample-2: $n_2 = 7$, $\Sigma(y_i - \bar{y})^2 = 95$

$$s_1^2 = \frac{1}{9-1}(175) = \frac{1}{8}(175) = 21.875$$

$$s_2^2 = \frac{1}{7-1}(95) = \frac{1}{6}(95) = 15.833$$

Here $s_1^2 > s_2^2$, $F = \frac{s_1^2}{s_2^2} = \frac{21.875}{15.833} = 1.3816$

$v_1 = n_1 - 1 = 9 - 1 = 8$

$v_2 = n_2 - 1 = 7 - 1 = 6$.

To verify the claim:

1) Null hypothesis (H_0): $\sigma_1^2 = \sigma_2^2$

2). Alternate hypothesis (H_1): $\sigma_1^2 \neq \sigma_2^2$ (two tailed)

3] Level of Significance : $\alpha = 5\% = 0.05$

$F_\alpha(v_1, v_2) = \cancel{2.98}\ 4.15$

4]. ~~critical value~~ $F = \dfrac{s_1^2}{s_2^2} = 1.3816$.

Test statistic

5]. $|F| = |1.3816| = 1.3816$

$F_\alpha(v_1, v_2) = 4.15$.

$|F| < F_\alpha(v_1, v_2)$

$\Rightarrow$ Ho is accepted.

$\Rightarrow$ There is no significant difference b/w variances.

$\Rightarrow$ They are drawn from normal populations with same variance.

<u>Prob</u>: Two random samples of sizes 9 and 6 gave the ~~random~~ following values of the variables:

Sample 1 : 15, 22, 28, 26, 18, 17, 29, 21, 24.

Sample 2 : 8, 12, 9, 16, 15, 10

Test the difference of estimates of population variances at 5% LOS.

Sol : $n_1 = 9$ $\quad n_2 = 6$.

$\bar{x} = (15+22+28+26+18+17+29+21+24)/9 = \dfrac{200}{9} = 22.222$

$\bar{y} = (8+12+9+16+15+10)/6 = 70/6 = 11.666$.

Computation table. $[\bar{x} = 22.222, \bar{y} = 11.666]$

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
15	8	-7.222	52.157	-3.666	13.439
22	12	-0.222	0.049	0.334	0.1115
28	9	5.778	33.385	-2.666	7.107
26	16	3.778	14.273	4.334	18.783
18	15	-4.222	17.825	3.334	11.115
17	10.	-5.222	27.269	-1.666	2.775
29		6.778	45.941		$\Sigma(y_i-\bar{y})^2 = 53.33$
21		-1.222	1.493		
24		1.778	3.161.		
			$\Sigma(x_i-\bar{x})^2 = 195.553$		

$$S_1^2 = \frac{1}{(n_1-1)} \Sigma(x_i-\bar{x})^2 = \frac{1}{8}(195.553) = 24.444$$

$$S_2^2 = \frac{1}{(n_2-1)} \Sigma(y_i-\bar{y})^2 = \frac{1}{6}(53.33) = 9.221.$$

Here $S_1^2 > S_2^2$.

$v_1 = n_1 - 1 = 9 - 1 = 8$

$v_2 = n_2 - 1 = 7 - 1 = 6.$

To verify claim:

1). $H_0: \sigma_1^2 = \sigma_2^2$

2) $H_1: \sigma_1^2 > \sigma_2^2$ (Right tailed)

3). LOS $\alpha = 5\%$

$F_v(v_1, v_2) = 4.82$

4] Test statistic is $F = \frac{S_1^2}{S_2^2} = \frac{24.444}{9.221} = 2.650$

5]. $|F| = 2.65$

$F_\alpha(V_1, V_2) = 4.82$

$|F| < F_\alpha(V_1, V_2)$

$\Rightarrow$ there is no significant difference b/w variance.

(Ho is accepted);

(chi-square)

χ^2 - Test for Population Variance:

Assumption: The parent population from which the random sample is drawn is normally distribution

This test is useful to check whether population variance (σ^2) has a specified value σ^2

That is to check whether $\boxed{\sigma^2 = \sigma_0^2}$

This is Ho

In this case,

the test statistic is

$$\chi^2 = \frac{ns^2}{\sigma^2} = \frac{\Sigma(x_i - \bar{x})^2}{\sigma^2} \text{ (or) } \frac{\Sigma(x_i - \bar{x})^2}{\sigma_0^2}$$

It follows χ^2- distribution with $V = n-1$ d.o.f.

Note: If χ^2 cal $< \chi^2_\alpha(V)$,

then Ho is accepted.

If χ^2 cal $> \chi^2_\alpha(V)$, then Ho is accepted.

If $\chi^2_{cal} > \chi^2_\alpha(v)$, then H_0 is rejected. Here $\chi^2_\alpha(v)$ is the critical value of χ^2 at α level of significance for $v = n-1$ degree of freedom.

Eg: A sample of 20 observation gave a standard deviation of 3.72. Is this compatable with the hypothesis that the sample is from a normal population with variance 4.35 = ? at 5% LOS..

Sol: Given,

Sample size is $n = 20$.

Sample standard deviation $s = 3.72$

Population Variance $\sigma^2 = 4.35$
$\downarrow$
σ_0^2.

To check the hypothesis that the sample is from a normal population with variance is 4.35.

That is we need to check whether $\sigma^2 = 4.35$.

So, we take $H_0: \sigma^2 = 4.35$.

The value of test statistic is $\chi^2 = \frac{ns^2}{\sigma^2}$

$$= \frac{20 \times (3.72)^2}{4.35}$$

$$= 63.624$$

It follows χ^2 distribution with $v = n-1 = 19$ degrees of freedom.

The critical value of χ^2 at $\alpha = 5\%$, $v = 19$ degrees of freedom is $\chi^2_\alpha(v) = 30.144$.

Here we have

$$\chi^2 \text{ cal} = 63.624, \quad \chi^2_\alpha(v) = 30.144.$$

$$\chi^2 \text{ cal} > \chi^2_\alpha(v).$$

$\rightarrow$ Null Hypothesis $H_0 : \sigma^2 = 4.35$ is rejected.

Errors in sampling:

Type - I error: If we reject the null hypothesis H_0 when it should be accepted, then we say that type - I has been made.

The probability of making Type - I error is denoted by α.

That is: P[Reject H_0 when it is true] $= \alpha$.

Type - II error: If we accept the NULL hypothesis H_0 when it should be rejected, then we say that Type - II error has been made.

The probability of macking Type - II error is denoted by β.

That is: P[Accept H_0 where it is wrong] $= \beta$.

Dischotomous Table:

Actual	Decision: Accept H_0	Reject H_0
H_0 is true	No error, correct decision prob $= 1-\alpha$	Type-I error wrong decision prob $= \alpha$
H_0 is wrong	Type-II error wrong decision. prob $= \beta$	No error. correct decision. prob $= 1-\beta$.

Prob: Weights in kgs of 10 students are given below 38, 40, 45, 53, 47, 43, 55, 48, 52, 49. Can we say that variance of distribution of weights of all students from which the above sample 10 students was drawn is equal to 20 kg? Test at 5%, 1% level of significance.

Sol: Take $H_0 : \sigma^2 = 20$.

Given,

sample size $n = 10$

sample mean $\bar{x} = \frac{\Sigma x_i}{n} = \frac{470}{10} = 47$

The variance $s^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2 = \frac{280}{10} = 28$

The value of Test statistic is $\chi^2 = \frac{ns^2}{\sigma^2}$

$= \frac{10(28)}{20}$

$= 14.$

degrees of freedom $v = n-1 = 9$.

Critical value of χ^2 for $v = 9$ degrees of freedom

is $= \begin{cases} 16.92 & \text{for } \alpha = 5\% \\ 21.67 & \text{for } \alpha = 1\% \end{cases}$

Here

$\chi^2_{cal} = 14$

$\chi^2_{\alpha}(v) = \begin{cases} 16.92 & \text{for } 5\% \\ 21.67 & \text{for } 1\% \end{cases}$

$\chi^2_{cal} < \chi^2_{\alpha}$ for both $\alpha = 5\%, 1\%$.

H_0 is accepted for both $\alpha = 5\%, 1\%$.