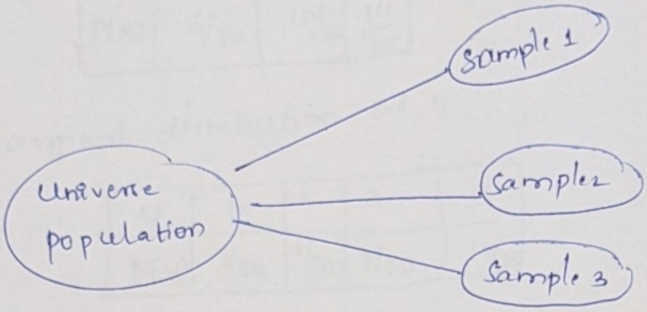
Chapter-4 Statistical Inference - 1



* Sampling:-

⇒ An aggregate of objects under study is called population or universe.

⇒ A finite subset of a universe is called as sample.

⇒ A sample is thus a small portion of the universe.

⇒ The number of individuals in a sample is called the sample size.

⇒ The process of selecting a sample from universe is called sampling.

* Parameters of statistics:-

The statistical constants of the population such as mean, variance, and proportion are known as the paramete

⇒ The population mean and variance are denoted by

mean = μ

variance = σ^2, standard deviation = σ

and that of sample, mean = $\bar{x}$

variance = $(\sigma_{\bar{x}})^2$ (or) s^2

and standard deviation = s (or) σ_x

Two types of sampling

(1) Cluster sampling:-

cluster sampling refers to type of sampling method with cluster sampling.

⇒ The researcher divides the population into seperate groups called clusters. then,

⇒ ~~syab~~ simple random sample of clusters is selected from the population.

(2) Stratified sampling:-

Divide the entire population into different subgroups called strata, ~~then~~ a simple random sample is drawn from each group.

* Sampling distribution of means:

Consider the population ~~for~~ which mean is 'μ' and Standard deviation is 'σ'.

suppose we draw a set of samples of certain size 'n'. For this population, we find the mean $\bar{x}$ for each of these samples.

the frequency distribution of these means is called a sampling distribution of means.

⇒ Let the mean and standard deviation of this distribution be $\bar{x}$ and $\sigma_{\bar{x}}$ respectively.

suppose the population is finite with size N_p, without replacement then $\bar{x}$ and s are related to μ and σ ~~through~~ the following ~~relations~~ formulae.

$$\boxed{\mu = \bar{x}}$$

$$\boxed{s = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}}$$

where 'n' = is the size of the sample.

⇒ Suppose the population is finite with size N_p, with replacement.

$$\boxed{\begin{array}{l}\mu = \bar{x} \\ S = \dfrac{\sigma}{\sqrt{n}}\end{array}}$$

ex:-

Q. A population contains 4 units → 3, 7, 11, 15. Obtain a sampling distribution of sample mean when samples are drawn of size '2'
(1) without replacement
(2) with replacement.

Soln:
(1) without replacement

3, 7, 11, 15

$N_p = 4$

$$\mu = \frac{\sum x_i}{N} = \frac{3+7+11+15}{4} = 9, \quad \boxed{\mu = 9}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} = \sqrt{\frac{36+4+4+36}{4}}$$

$$\boxed{\sigma = 2\sqrt{5} = 4.47}$$

$$\boxed{\sigma^2 = 20}$$

Sample WOR:-

(3, 7), (3, 11), (3, 15), (7, 11), (7, 15), (11, 15).

$$\boxed{\text{mean} \Rightarrow 5, 7, 9, 9, 11, 13}$$

(1) $\frac{3+7}{2} = 5$

(2) $\frac{3+11}{2} = 7$

(3) $\frac{3+15}{2} = 9$

mean of each sample } = 5, 7, 9, 9, 11, 13.

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{5+7+9+9+11+13}{6} = 9$

$\bar{x} = 9 =$

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

$\Rightarrow \frac{1}{6} \left((5-9)^2 + (7-9)^2 + (9-9)^2 + (9-9)^2 + (11-9)^2 + (13-9)^2 \right)$

$= \frac{1}{6} (16 + 4 + 4 + 16)$

$\boxed{S^2 = 6.667}$

$\boxed{S = 2.58}$

verification:

$S = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{Np-n}{Np-1}}$

$= \frac{4.47}{\sqrt{2}} \sqrt{\frac{4-2}{4-1}}$ (n = size = 2)

$= \frac{4.47}{\sqrt{2}} \sqrt{\frac{2}{3}}$

$= \frac{4.47}{\sqrt{3}} \Rightarrow 2.58$ //

with replacement:

sample WR: (3,7), (3,11), (3,15), (3,3), (7,3), (7,11),
(7,7), (7,15), (11,3), (11,7), (11,11), (11,15)
(15,7), (15,3), (15,11), (15,15).

$x_i \rightarrow$ 5, 7, 9, 3, 5, 9, 7, 11, 7, 9, 11, 13, 11,
9, 13, 15.

Mean of each sampling $\left.\begin{matrix} \\ x_i \end{matrix}\right\}$ = 5, 7, 9, 3, 5, 9, 7, 11, 7, 9, 11, 13, 11, 9, 13, 15.

$$\bar{x} = \frac{144}{16}$$

$$\boxed{\bar{x} = 9}$$

$$S = \frac{\sigma}{\sqrt{n}}$$

$$S^2 = \frac{1}{16}\sum(x_i-\bar{x})^2$$

$$= \frac{1}{16}\left[(5-9)^2 + (7-9)^2 + \cancel{(9-9)^2} + (3-9)^2 + (5-9)^2 + \cancel{(9-9)^2} + (7-9)^2 + (11-9)^2 + (7-9)^2 + \cancel{(9-9)^2} + (11-9)^2 + (13-9)^2 + \cancel{(9-9)^2} + (13-9)^2 + (15-9)^2\right].$$

$$= \frac{1}{16}(16 + 4 + 36 + 16 + 4 + 4 + 4 + 4 + 16 + 16 + 36)$$

$$S^2 = \frac{160}{16} = 10 \qquad \boxed{S = 3.16}$$

Verification:-

$$S = \frac{\sigma}{\sqrt{n}} = \frac{4.47}{\sqrt{2}} = 3.16$$

* **Standard error:- (S.E)**

The standard deviation of the sampling distribution of a statistic is known as "Standard error".

→ It plays an important role in the theory of Large samples ($n \geq 30$).

→ It forms basis of testing of hypothesis.

Statistic	S.E	
1) $\bar{x}$	$\sigma/\sqrt{n}$	
2) S	$\sqrt{\sigma^2/2n}$	
3) proportion (p)	$\sqrt{\frac{pq}{n}}$	$(q = 1-p)$
4) difference of two sample means $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	
5) $S_1 - S_2$ (dist of two S.Ds)	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$	
6) $P_1 - P_2$ (diff between two proportions)	$\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$	

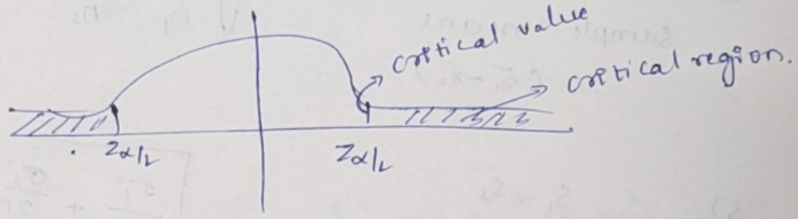
Null hypothesis:- A definite statement about the population parameter called Null hypothesis. It is denoted by H_0.

Alternate hypothesis:- Any hypothesis which is complementary to Null hypothesis (H_0) is called alternate hypothesis. It is denoted by H_1.

Type I error: we reject the hypothesis 'H' when it should be accepted (true one is rejected).

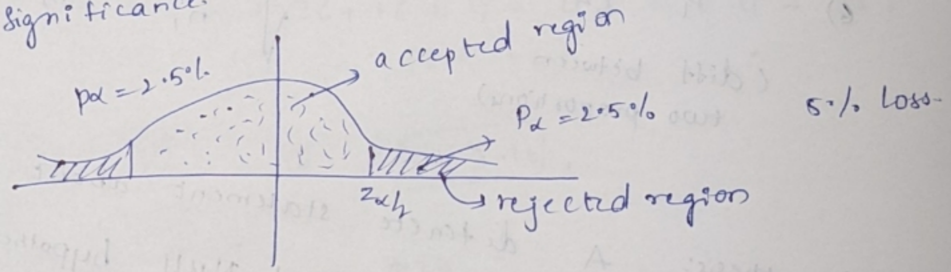
Type II error: we accept the hypothesis 'H' when it should be rejected (False one is accepted).

Critical region: A region in sample space 'S' which amounts to rejection of H_0 is termed as critical region or rejection.



Level of significance: (LOS)

The probability 'α' that the random value of statistics 't' belongs to region is known as level of significance.



Confidence interval:

Interval estimation which gives the end points between which the parameters of the population lies.

Confidence Limits for mean μ is given by

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

mean LOS standard error for mean

Maximum error to estimate is given by

$$\boxed{E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

Size of sample, $n = \left(Z_{\alpha/2} \frac{\sigma}{E}\right)^2$.

Note:- for 95% C.I $Z_{\alpha/2} = 1.96$

for 99% C.I $Z_{\alpha/2} = 2.58$

Confidence Interval (C.I) for proportion P.

$$p \pm Z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

↳ standard error.

point estimation

for mean $\bar{x}$ ⇒, $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

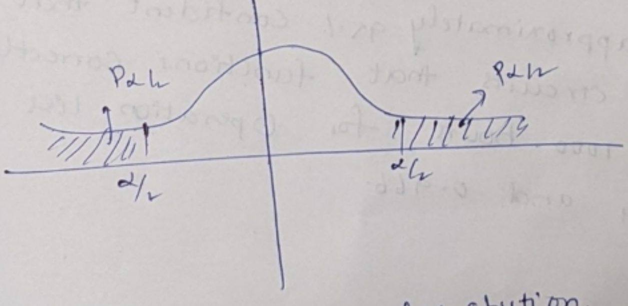
for difference of proportion, $(P_1 - P_2) \pm Z_{\alpha/2} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$

for difference of means $\bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$n \geq 30$ → large sample, use normal distribution

$n \not\geq 30$ → small sample, use t- distribution.

when σ is unknown, use t- distribution.



two tailed test.

Spread is less in t-distribution.

Spread is more in normal distribution.

⑦ $n = ?$

error $= \pm 10{,}000$

C.I $= 95\%$

$\sigma = 30{,}000$

$$n = \left(Z_{\alpha/2}\,\frac{\sigma}{E}\right)^2 = \left(1.96 \times \frac{\overset{3}{\cancel{3000}}}{\cancel{10000}}\right)^2 = 34.57 \simeq 35$$

⑧ C.I $= 95\%$

$\sigma = 2$ light years

mean $= d$ light years

error $= \pm 0.5$

$$n = \left(Z_{\alpha/2}\,\frac{\sigma}{E}\right)^2 = \left(1.96 \times \frac{2}{0.5}\right)^2 = 61.46 \simeq 62.$$

⑨ $p =$ proportion $= 0.91$

$q = 1 - 0.91 = 0.09$

$n = 100$

for 95%

C.I, $p \pm 1.96\sqrt{\frac{pq}{n}}$ $\quad 0.91 + 1.96\sqrt{\frac{0.91 \times 0.09}{100}}$

C.I (0.91 ± 0.056)

$(0.966,\ 0.854)$

we can get approximately 95% confident that true proportion of circuits that functions correctly during first 1000 hours " for operation. lies between 0.854 and 0.966.

for 99%

C.I $\quad p \pm 2.58\sqrt{\frac{pq}{n}}$ $\quad p \pm 2.58\sqrt{\frac{0.91 \times 0.09}{100}}$

C.I (0.91 ± 0.077)

C.I $(0.987,\ 0.836)$

for 96% C.I $Z_{\alpha/2} = 1.76$ (from table)

(10) This is proportion, $n = 500$

$p = \frac{60}{500} = 0.12$

$q = 0.88$

C.I $\Rightarrow$ $p \pm 1.76 \sqrt{\frac{pq}{n}}$

$= 0.12 \pm 1.76 \sqrt{\frac{0.12 \times 0.88}{500}}$

$\Rightarrow 0.12 \pm 0.0255$

C.I $(0.145, 0.095)$

(11) $n = 100$,

$\bar{x} = 67.45$

$s = 2.92$ (SD of sample)

for 97% $Z_{\alpha/2} = 1.89$

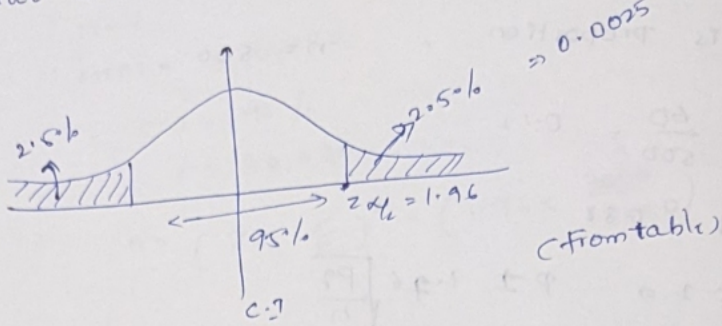
C.I for mean, $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

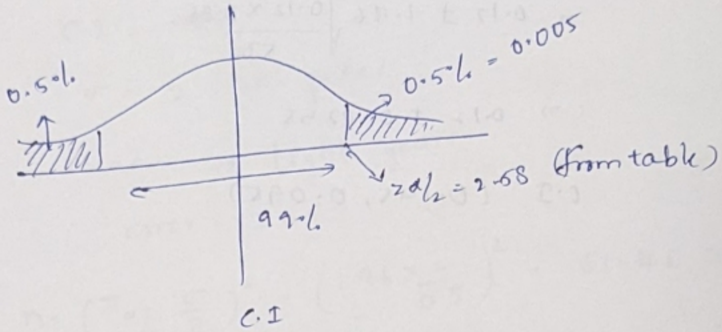
$67.45 \pm (1.89) \times \frac{2.92}{\sqrt{100}}$

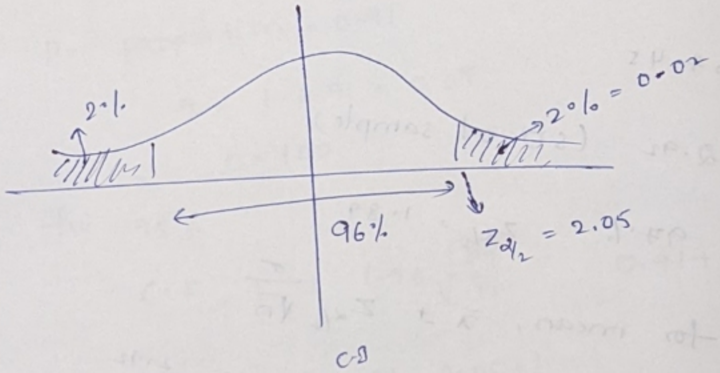
C.I (67.45 ± 0.551)

C.I $(68.001, 66.9)$

for two tail







LP's

13

$P_1 - P_2 = 0.589 - 0.619 \Rightarrow -0.03$

Construct 95% C.I for this difference.

C-I, Confidence Intervals

Soln:

C.I for difference of proportion is,

$$\left((p_1 - p_2) \pm z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right) \quad \text{——} \quad (1)$$

where, $p_1 = 0.589$ $\quad q_1 = 1 - p_1$

$p_2 = 0.619$ $\quad q_2 = 1 - p_2$

$p_1 - p_2 = -0.03$

$z_{\alpha/2} = 1.96$ for 95% C.I.

$n_1 = n_2 = 375$.

From (1)

$$(0.589 - 0.619) \pm 1.96\sqrt{\frac{0.589(1-0.589)}{375} + \frac{0.619(1-0.619)}{375}}$$

$$\Rightarrow -0.03 \pm 1.96\left(\sqrt{\frac{0.2421}{375} + \frac{0.236}{375}}\right)$$

$$\Rightarrow -0.03 \pm 1.96\left(\sqrt{0.00064 + 0.000629}\right) \approx -0.0~~0.0356~~$$

$$\Rightarrow -0.03 \pm \cancel{0.0508}\ 0.07.$$

$$\Rightarrow (0.04, -0.1)$$

i.e., we are 95% confident that true difference in proportion lies in the interval $(-0.1, 0.04)$ and note that since this interval contains the number '0', it is possible that there is really no difference in two population proportion P_1 and P_2.

AP:-

(14) Given, $n_1 = 500$, $n_2 = 450$ (size of samples)

$p_1 = \frac{178}{500} = 0.356$, $p_2 = \frac{200}{450} = 0.444$.

$q_1 = 1 - p_1$
$= 1 - 0.356$
$\boxed{q_1 = 0.644}$

$q_2 = 1 - p_2$
$= 1 - 0.444$
$\boxed{q_2 = 0.556}$

$$\Rightarrow (P_1-P_2) \pm 1.96\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$

$$\Rightarrow (-0.088) \pm 1.96\sqrt{\frac{0.356(0.644)}{500} + \frac{0.444(0.556)}{450}}$$

$$(-0.088) \pm 1.96\left(\sqrt{0.000458 + 0.000548}\right)$$

0.001006

$$-0.088 \pm 1.96\,(0.03174)$$

$$\Rightarrow -0.088 \pm 0.0622$$

$$\Rightarrow (\cancel{-0.045, -0.131})$$

$$\Rightarrow (-0.0258, -0.1502)$$

LP:-

15

$n_1 = 250$	$n_2 = 400$
$\bar{x}_1 = 120$	$\bar{x}_2 = 124$
$s_1 = 12$	$s_2 = 14$

and C.I for difference of means

$$\Rightarrow (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\Rightarrow (120-124) \pm 2.58\left(\sqrt{\frac{144}{250} + \frac{196}{400}}\right)$$

$$= \cancel{-4 \pm 2.58}\sqrt{}$$

$$= -4 \pm 2.58\,(1.03247)$$

$$\Rightarrow -4 \pm 2.6637$$

$$= (-1.3363, -6.6637)$$

* **t-Test:-**

when the size of sample is less than 30, then the sample is called samall sample.

This t- distribution is used when sample size is less than 30, and the population standard deviation is unknown.

→ The quantity, $\nu = n-1$, is called the number of degrees of freedom.

→ (The degrees of freedom refer to the number of pieces of independent information used to estimate of variancy.

→ we can define the confidence limits for the t-distribution for the population means, the confidence limit are given by $\left(\bar{x} \pm t_{\alpha/2} \left(s/\sqrt{n}\right)\right)$.

where, $\pm t_{\alpha/2}$ are the critical values or confidence ~~quotient~~ coefficients whose values depend on the level of significance desire and sample size.

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

LP :-

(12) Given,

$n = 10$

578, 572, 570, 568, 572, 571, 570, 572, 596, 548.

Find 99% C.I for the mean.

Soln: Since $n = 10 \Rightarrow$ small sample.

use t-distribution

C.I for mean

$$\left(\bar{x} \pm t_{\alpha/2} \left(s/\sqrt{n}\right)\right).$$

$\bar{x} = \frac{1}{n} \Sigma x_i$

$= \frac{1}{10} (578 + 572 + \cdots 548 \}$

$\boxed{\bar{x} = 571.7}$

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

$= \frac{1}{10-1} \{ (578 - 571.7)^2 + (572 - 571.7)^2 + \cdots + (548 - 571.7)^2 \}$.

$= \frac{1}{9} (1212.10)$

$= \cancel{136}$

$s^2 = 134.67$

$\boxed{s = 11.6}$

tabulated $t_{0.01}$ (99% C.I) with degrees of freedom is $(10-1) = 9$ is,

$t_{0.01} = 3.25$

$C.I = \left(\bar{x} \pm t_{\alpha/2} \left(s/\sqrt{n} \right) \right)$

$\Rightarrow \left(571.7 \pm 3.25 \left(11.6/\sqrt{10} \right) \right)$

$\left(571.7 \pm 3.25 (3.668) \right)$

$\rightarrow (583.621, 559.779)$ ⇒ The mean samples lies in between them.

* Procedure for testing of Hypothesis:-

1) Null hypothesis: Set up the null hypothesis, H_0

2) Alternative hypothesis:- set up the Alternative hypothesis, H_1.

This will enable us to decide whether we have to use single tailed or 2-tailed test.

3) Level of significance:- (LOS)

Choose appropriate LOS α

4) Test statistic :- Compute the test statistic,

$$\boxed{Z_{cal} = \left| \frac{t - E(t)}{S.E(t)} \right|} = \frac{\text{observed-expected}}{S.E.}$$

5) conclusion:-

Compare the computed p value of Z_{cal} with the significant value P_α (tabulated value) at the given LOS.

$\Rightarrow P_{cal} < P_\alpha$ (calculated p-value < tabulated p-value)

then reject H_0. (Null hypothesis).
(accepted H_1)

$\Rightarrow P_{cal} > P_\alpha \Rightarrow$ accept H_0 (rejected H_1)

$\Rightarrow$ Test of Significance for single mean and proportion:-

Proportion $\Rightarrow$ $$Z_{cal} = \left| \frac{p - P}{\sqrt{PQ/n}} \right|$$

$p \rightarrow$ proportion
$P \rightarrow$ probability of population.

$$\therefore Z_{cal} = \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right|$$

Difference of proportion:

$$Z_{cal} = \left| \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \quad \text{where } \boxed{P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}}.$$

Difference of means:-

$$Z_{cal} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$$

For Large sample use Normal distribution

For small sample use t- distribution, with dof (n-1)

I.P:-

(18)

$\mu = 578$ (given)

$\left.\begin{array}{l}\bar{x} = 571.7 \\ S = 11.6\end{array}\right\}$ calculated in previous question (12).

$H_0: \bar{x} = \mu$ (or $\mu = 578$

$\mu \neq 578$.

$H_1: \bar{x} \neq \mu$ (two tailed test)

$t_\alpha = t_{0.05}$ (dof = 9)

$t_\alpha = 2.262$

Since it is a small sample,

$$t_{cal} = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right|$$

$$= \left| \frac{571.7 - 578}{11.6/\sqrt{10}} \right|$$

$t_{cal} = 1.717$

$t_{cal} < t_2$ $(1.717 < 2.262)$

$\Rightarrow H_0$ is accepted (Reverse of p-values).

i.e $\bar{x} = \mu$ means there is no significance difference between mean of the sample and mean of the population at 5% LOS.

LP:-

(16) Given, $n = 100$ (big sample, $\Rightarrow$ use Normal distribution).

$\bar{x} = 1570$ $\mu = 1600$.

$s = 120$

Soln: $H_0: \bar{x} = \mu$

$H_1 \Rightarrow \bar{x} \neq \mu$ (Two tailed test).

$Z_\alpha = 1.96$ at 5% LOS i.e $P_\alpha = 0.05$

$$Z_{cal} = \left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right|$$

$$= \left|\frac{1570 - 1600}{120/\sqrt{100}}\right|$$

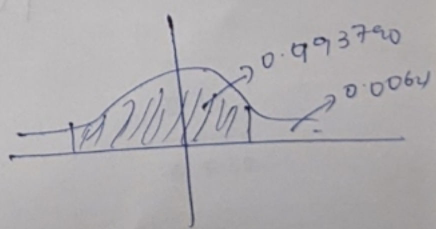
$Z_{cal} = 2.5$

$Z_{cal} > Z_\alpha$ using Z value H_0 is rejected at 5% LOS.

$\phi(z_{cal}) = \phi(2.5) = 0.993790$

$1 - \phi(2.5) = 1 - 0.993790$

$= 0.00621$



Since it is '2' tail,

$P_{cal} = 2(0.00621)$

$P_{cal} = 0.01242$

$P_{cal} = 0.0124$

$P_\alpha = 0.05$ (tabulated values)

$P_{cal} < P_\alpha \Rightarrow H_0$ is rejected

$\Rightarrow$ mean of the sample and mean of population are significantly different at 5% LOS

LP:-

(17) $n_1 = 40,\ n_2 = 50$

$\bar{x}_1 = 74 \quad s_1 = 8$

$\bar{x}_2 = 78 \quad s_2 = 7$

Soln:-

$H_0: \bar{x}_1 = \bar{x}_2$ (same)

$H_1: \bar{x}_1 \neq \bar{x}_2$ (not same)

$Z_\alpha = 1.96$ at 5% LOS $\quad P_\alpha = 0.05$

$Z_\alpha = 2.58$ at 1% LOS $\quad P_\alpha = 0.01$

$$Z_{cal} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

$$= \left| \frac{74-78}{\sqrt{\frac{64}{40} + \frac{49}{50}}} \right| = \left| \frac{4}{1.606} \right|$$

$$= 2.490$$

$$\boxed{Z_{cal} = 2.49}$$

(area) $\phi(Z_{cal}) \Rightarrow \phi(2.49) \Rightarrow 0.993613$

$1 - \phi(2.49) \Rightarrow 1 - 0.993613$

$\Rightarrow 0.006387$

$P_{cal} = 2\,(0.006387)$

$\boxed{P_{cal} = 0.012774}$

$P_{cal} < P_\alpha$ (5% LOS) $\Rightarrow$ H_0 is rejected.

$P_{cal} > P_\alpha$ (1% LOS) $\Rightarrow$ H_0 is accepted.

(20) Given,

$P_1 = \frac{300}{400}$ (university)

$P_2 = \frac{300}{500}$ (colleges)

H_0 is $\Rightarrow$ $P_1 = P_2$

H_1 $\Rightarrow$ $P_1 \ngtr P_2$ (one tailed test)

consider $P_\alpha = 0.05$ (at 5% LOS)

$$Z_{cal} = \left| \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$

$$P \Rightarrow \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2} \Rightarrow \frac{\frac{300}{400}(400) + \frac{300}{500}(500)}{900}$$

$\boxed{P = 0.666}$

$Q = 1 - P$

$= 1 - 0.666$

$= 0.34$

$$Z_{cal} = \left| \frac{0.75 - 0.6}{\sqrt{0.66(0.34)\left(\frac{1}{400} + \frac{1}{500}\right)}} \right| = \frac{0.15}{0.032}$$

$\boxed{Z_{cal} = 4.68}$

area, $\phi(Z_{cal}) \simeq 1$,

when $Z_{cal} > 3.9$ then (H_0 is rejected)

Note:-

If sample is small use t- distribution and Compare the value of t_{cal} and tabulated t-value

(~~If~~ ~~can~~) ⇒ if $t_{cal} < t_{tab}$ ⇒ accept H_0

⇒ if ques contains less than or greater than or increasing decreasing then use one tail test.

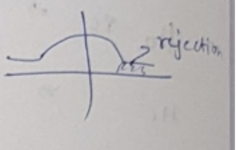
and p-value should not be multiplied by '2' for one tail test.

Note:

1) If $n \not< 30$ and S.D is unknown then use t- distribution.

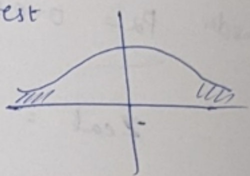
If $n \not> 30$ and S.D is known then use z- distribution.

2) calculate Z_{cal} by Test statistic.



$P_{cal} = 1 - \Phi(z_{cal})$ → one tail test

$P_{cal} = 2(1 - \Phi(z_{cal}))$ → Two tail test



If $P_{cal} < P_\alpha$ i.e observed prob < expected prob

H_0 is Rejected.

~~If~~ $P_{cal} > P_\alpha$ ⇒ H_0 is accepted.

Q:- A random sample of 100 recorded deaths in the united states during the past year showed an average life span of 71.8 years, assuming a population S.D is 8.9 years. Does this seem to indicate that the mean life span today is greater than 70 years.

Soln:

$n = 100$ $\mu = 70$

$\bar{x} = 71.8$ (mean)

$S.D = 8.9$ (σ)

Null hypothesis, $H_0 : \mu = 70$

Alternate hypothesis $H_1 : \mu > 70$ (one tail).

Consider, $P_\alpha = 0.05$

test statistic for mean

$$Z_{cal} = \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| \Rightarrow \left| \frac{71.8 - 70}{8.9/\sqrt{100}} \right|$$

$$\Rightarrow \left| \frac{1.8 \times 10}{8.9} \right|$$

$$\Rightarrow \boxed{Z_{cal} = 2.0225}$$

$P_{cal} = 1 - \Phi(2.02)$

$= 1 - 0.9783$

$\boxed{P_{cal} = 0.0217}$

$P_{cal} < P_\alpha \Rightarrow H_0$ is rejected

H_1 is accepted.

mean lifespan is greater than 70 and 5% level of significance.

Q: According to Chemical engineering an important property is fiber is water absorbacy the average percent absorbacy of 35% randomly selected 35 pieces found to be 20 with S.D of 1.5.
The Random sample is 35 pieces of acceterate heated at average percent of 12 with the S.D of 1.25.

Is there strong evidence that the population mean percent absorbacy of cotton fibre is significantly higher than the mean for acceterate.
Assume that the percent absorbacy is approx. normally distributed

Sol:

$n_1 = 35$ $\quad$ $n_2 = 35$

$\bar{x}_1 = 20$ $\quad$ $\bar{x}_2 = 12$

$S_1 = 1.5$ $\quad$ $S_2 = 1.25$

Null hypothesis : $H_0, \bar{x}_1 = \bar{x}_2$

$H_1 \Rightarrow \bar{x}_1 > \bar{x}_2$ (one tailed)

$P_\alpha = 0.01$

Test statistic for diff of means.

$$Z_{cal} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$$

$$\Rightarrow \left| \frac{20-12}{\sqrt{\frac{3.8125}{35}}} \right|$$

$$\Rightarrow \left| \frac{20-12}{0.3300} \right|$$

$\Rightarrow 24.24$

Since $Z_{cal} > 3.9 \Rightarrow$ there is no rejection region.

$\Rightarrow H_0$ is rejected

$\Rightarrow H_1$ is accepted

Q: 10 individuals are chosen at random from a population and their height in inches are found to be 63, 63, 68, 67, 68, 69, 70, 70, 71, 71. test the hypothesis that mean height of universe is 66 inches.

Soln: $\mu = 66$

$n = 10$, S.D is unknown.

use t-distribution.

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

$$\Rightarrow \frac{1}{10-1} (2(63-67.8)^2 + 2(68-67.8)^2 + (67-67.8)^2 + 2(70-67.8)^2 + 2(71-67.8)^2)$$

$$\Rightarrow \frac{1}{9}(46.08 + 2(0.04) + 0.64 + 2(4.84) + 2(10.24))$$

$\Rightarrow \frac{1}{9}(76.96)$

$s^2 = 8.55$

$\boxed{s = 2.924}$.

$s^2 = 9.06$

$\boxed{s = 3}$

$H_0 \Rightarrow \mu = 66$

$H_1 \Rightarrow \mu \neq 66$

$t_{0.05}(9) = 2.262 = t_2$

$P_\alpha = 0.05$

$P_{cal} = 0.075$

$P_\alpha = 0.05$

$P_{cal} > P_\alpha$

H_0 is acepted.

test statistics:-

$$t_{cal} \Rightarrow \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| \Rightarrow \left| \frac{67.8 - 66}{3/\sqrt{10}} \right| \Rightarrow \frac{1.8\sqrt{10}}{3}$$

$\boxed{t_{cal} = 1.897}$

$$P = \frac{0.10 + 0.05}{2}$$

$$\Rightarrow \frac{0.15}{2}$$

$P = 0.075$

Q(21):

$P_1 = \frac{8}{100}$, $P_2 = \frac{12}{200}$

$H_0 : P_1 = P_2$

$H_1 : P_1 > P_2$