18/12/2020.

UNIT 2: TESTING OF HYPOTHESIS

Hypothesis is a statement regarding the population parameter.

Hypothesis is of two types

1) Null Hypothesis
2) Alternate Hypothesis

Null Hypothesis: It is a no difference statement.

It is denoted by H_0.

ex: $H_0 : \mu = 3$

ex: $H_0 : \mu_1 = \mu_2$

ex: $H_0 : \sigma^2 = 2$

Alternative Hypothesis: It is complimentory to Null hypothesis. It is denoted by H_1.

ex: $H_1 \mu_1 \neq \mu_2$ ex: $H_1 : \mu \rightarrow 2$.

$H_1 \sigma_1 \neq \sigma_2$

Alternative hypothesis is of two types:

1) Two tailed alternative
2) Single tailed alternative.

→ Two Tailed Alternative:

ex: $H_0 : \mu_1 = \mu_2$ ex: $H_0 : \sigma^2 = 30$

$H_1 : \mu_1 \neq \mu_2$ $H_1 : \sigma^2 \neq 30$

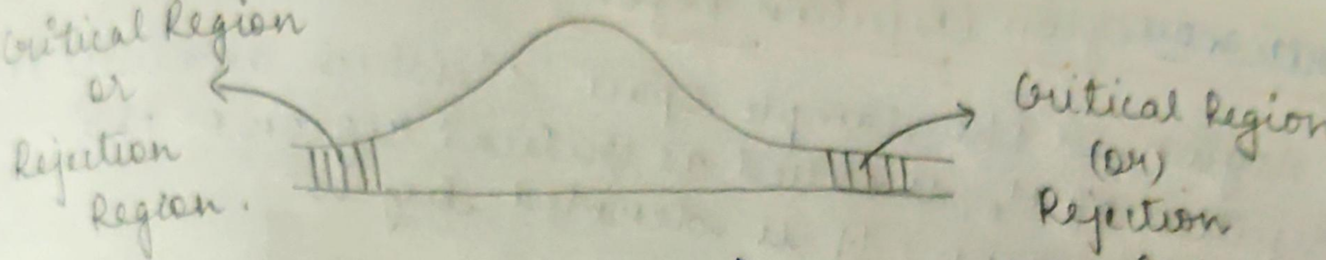


fig: Two Tailed Alternative

Single Tailed Alternative:

H is divided into two types:

1) Right Tailed Alternative
2) Left Tailed Alternative.

→ Right Tailed Alternative.

ex: $H_0 : H_1 = H_2$ ex: $\sigma^2 = 30$

$H_1 : H_1 > H_2$ $\sigma^2 > 30$

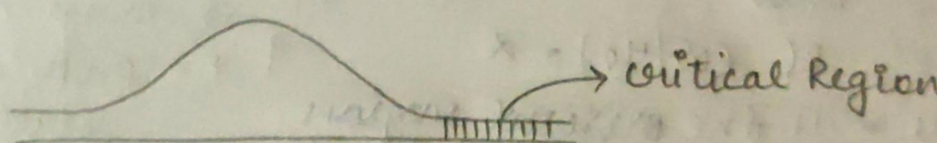


fig: Right Tailed Alternative

→ Left Tailed Alternative.

ex: $H_0 : H_1 = H_2$ ex: $H_0 : \sigma^2 = 30$

$H_1 : H_1 < H_2$ $H_1 : \sigma^2 < 30$

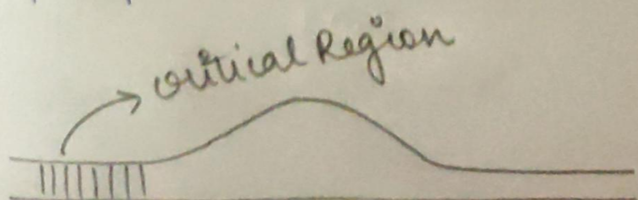
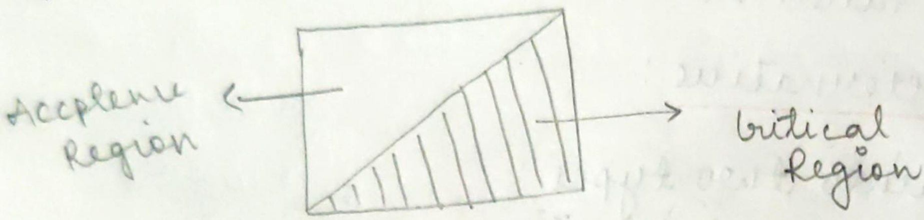


fig: Left Tailed Alternative.

Critical Region (Rejection Region)

A region in the sample space 's' which amounts to rejection of H_0, is called as critical region or the rejection region. It is denoted by w.



Sample Space (S)

27/12/2020.

Level of Significance:

If w is the critical region, and $t = t(x_1, x_2 \ldots x_n)$ is the value of the stastic based on random sample sized of size 'n'. then,

$$P(t \in w \mid H_0) = \alpha$$

w - omega - crtical region
t - probality of sample stastic.

Here α is the ~~critical region~~ level of significance

and $P(t \in \overline{w} \mid H_1) = \beta$

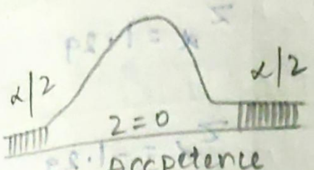
α - level of significance

$$w \cup \overline{w} = S$$
$$w \cap \overline{w} = \phi.$$

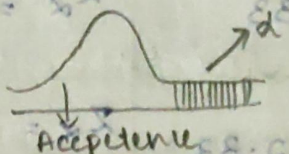
Critical Value or Significant value:

The value of test statistic which seprates the critical region and the accpetence region is called as critical value or significant value.

It depends on
1) Level of significance
2) Alternative hypothesis



Two tailed alternative



Right Tailed alternative.

Here the values α/2, α/2 is dividing the critical region with the accpetence region so it is called as critical value.

Types of errors:

There are two types of errors
(1) Type I error
(2) Type II error

Type I Error:

Probablity of rejeting a good lot, is called as a type I error.

P(rejeting a good lot) = α

Type II Error:

Probablity of accpeting a bad lot, is called as an type II ~~test~~ error.

P(accpecting a bad lot) = β

This is called as consumer's risk.

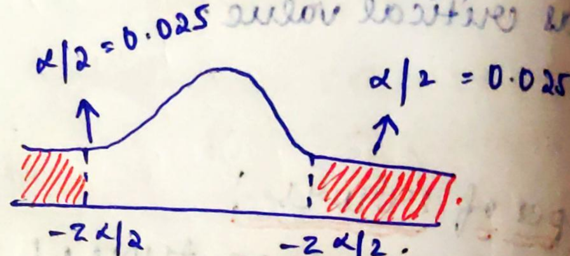
	1.1.%	5.%	10.%
Two tailed test	$z_{\alpha/2} = 2.58$	$z_{\alpha/2} = 1.98$	$z_{\alpha/2} = 1.65$
Right tailed test	$z_{\alpha} = 2.33$	$z_{\alpha} = 1.65$	$z_{\alpha} = 1.29$
Left tailed test.	$z_{\alpha} = -2.33$	$z_{\alpha} = 1.65$	$z_{\alpha} = -1.29$

Power of test : $1-\beta$ is called power of test.

5.% Two Tailed.

$\alpha = 0.05 = 5\%$

$\alpha/2 = \dfrac{0.05}{2} = 0.25$



Total area under normal curve = 1.

From $-\alpha/2$ to z_α area is

$$\begin{array}{r} 1.000 \\ -\ 0.025 \\ \hline 0.97 \end{array}$$

a particular, set of values $x_1, x_2, x_3 \dots x_n$.
ien sample $x_1, x_2 \dots x_n$ L becomes a function
ble θ, the parameter.

Procedure of Testing Hypothesis:

1) Setup up H_0
2) Setup H_1
3) Choose the level of significance, α
4) Calculate the test statistics under H_0
5) compare the calculated value of test statistics with table value.

→ For two tailed & right test

Calculated value < Table value, Accept H_0
Calculated value > Table value, Reject H_0

→ left tailed test

calculated value < Table value, Reject H_0
Calculated value > Table value, Accept H_0

08/01/2021

Test for single Proportion:

1)

$H_0: P = 30$	$H_0: P = 30$	$H_0: P = 30$
$H_1: P \neq 30$	$H_1: P > 30$	$H_1: P < 30$

$$\boxed{z = \frac{x - np}{\sqrt{npq}} \sim N(0,1)}$$

n = no. of times a coin is tossed.
p = prob of sucess
q = prob of failure.

2) $$\boxed{Z = \frac{\hat{p} - P}{\sqrt{PQ/n}} \sim N(0,1) \quad \hat{p} = \text{sample proportion},\ P = \text{Population proportion},\ Q = 1 - P,\ n = \text{Sample size}}$$

3) Confidence interval for true proportion

$$\boxed{\hat{p} - Z_{\alpha/2}\sqrt{pq/n} < P < \hat{p} + Z_{\alpha/2}\sqrt{pq/n}}$$

example:

1) A coin is tossed 960 times and head turned up 183 times. Is the coin biased?

sol^n:

H_0: coin is unbaised i.e., Prob. getting heads = Prob getting tails

H_1: coin is biased :

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0,1)$$

n = no. of times coin is tossed = 960

X = no. of success = 183

p = prob. of success = 0.5

q = prob. of failure = 0.5

$$Z = \frac{183 - (960 \times 0.5)}{\sqrt{960 \times 0.5 \times 0.5}} = \cancel{\frac{87840}{15.4919}} = -19.17126$$

articular, set of values $x_1, x_2, x_3 \cdots x_n$
sample $x_1, x_2 \cdots x_n$ L becomes a function

Table value of z at 5% level significance for two tailed alternative test is 1.96

calculated value > Table value

Reject H_0

Inference: coin is biased.

2) In a consignment of oranges, a random sample of 64 reveled that 14, oranges were bad. Is it resonable to ensure that 20%, of the oranges are bad?

Sol: $H_0 : P = 20\% = \frac{20}{100} = 0.2$

$H_1 = P \neq 0.2$

$$z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

x = no. of bad oranges = 14
n = sample size = 64
P = Population prop = 0.2
$Q = 1 - P = 1 - 0.2 = 0.8$
p = sample proportion $= \frac{x}{n} = \frac{14}{64} = 0.2187$

$$z = \frac{0.2187 - 0.2}{\sqrt{0.2 \times 0.8 / 64}}$$

$= 0.374$

Table value at z at 5% level of signifanie for a two tailed test is 1.96

calculated value < Table value.

Accpect H_0.

Inference : In the consignment 20% of apples are bad.

3) In a study designed to investigate wheather ~~some~~ certain detonators in coal mining meet the requirment that atleast 90% will ignite the explosive when charged it is found that 174 of 200 detonators function ~~property~~ properly. Test the null hypothesis $P = 0.9$ against the alternative hypothesis $P < 0.9$ at 5% and 1% level of significance.

$H_0 : P = 0.9$

$H_1 : P < 0.9$

$$Z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

$= x = 174$, $n = 2020$ $P = 0.9$

$Q = 1 - P = 1 - 0.9 = 0.1$

$$p = \frac{x}{n} = \frac{174}{200} = 0.87$$

$$Z = \frac{0.87 - 0.9}{\sqrt{(0.9)(0.1)/200}} = -1.4142$$

Table value of Z, at 5% loss for a left tailed test is -1.65

calculated value > Table value.

Accpect H_0.

Table value of Z at 1% loss for a left tailed test is -2.33.

calculated value > Table value ; Accpect H_0.

Inference : P = 0.9 is true at 5% and 1% levels of significince.

4) 20 people were attacked and only 18 survived. will you reject the hypothesis the survial rate attacked by this ~~diseasease~~ disease is 85% favour of the hypothesis that is more 5% level of significance.

Sol): $H_0 = 0.85 = P$

$H_1 : P > 0.85$

$$z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

$x = 18 \quad n = 20 \quad P = 0.85$

$Q = 1 - P = 1 - 0.85 = 0.15$

$$p = \frac{x}{n} = \frac{18}{20} = 0.9$$

$$Z = \frac{0.9 - 0.85}{\sqrt{\frac{0.85 \times 0.15}{20}}} = 0.6262$$

Total value of z at 5% los for a right tailed test is 1.65

Inference, cal val < Table value.

P = 0.85 is true, the survial rate is 0.85

CONFIDENCE INTERVAL PROBLEM FOR TRUE PROPORTIONS.

Note points:

* Confidence interval problems are always Two-tailed.
* Confidence interval is the same as confidence limits.

↳ Both have upper and lower bounds.

PROBLEMS:

① In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 99% confidence interval for the corresponding true proportion.

Soln: Given that:

$n = 160$

$x = 24$

$p = x/n = \frac{24}{160} = 0.15$

$q = 1-p = 1-0.15 = 0.85$

Given confidence region → 99

Critical region → 100 − 99 = 1

⇒ $\alpha = 1\%$ level of significance. Since it is two-tailed, we consider $\alpha/2 = \frac{0.01}{2} = 0.005$

$\Rightarrow 100 - 0.005$

$\Rightarrow 99.995$

[Why are we doing this? It is two tailed. The fraction part is considered as 0.005 → we take the absolute value disregarding the sign. Knowing the value for any one of the extreme ends is sufficient. Hence, we redefine the confidence level].

We consider the fraction part i.e. $\boxed{0.995}$

Search this value in the table (Normal Distribution table).

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	$\boxed{0.08}$	0.09
0.0										
0.1										
0.2										
⋮										
$\boxed{2.5}$									0.995	

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

$$p - Z_{\alpha/2}\sqrt{pq/n} < P < p + Z_{\alpha/2}\sqrt{pq/n}$$

$$0.07717 < P < 0.2228.$$

② A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Obtain 98% confidence limits.

Soln: Given:

$n = 500$

$x = 60$

$p = 0.12$

$q = 1 - p = 0.88$

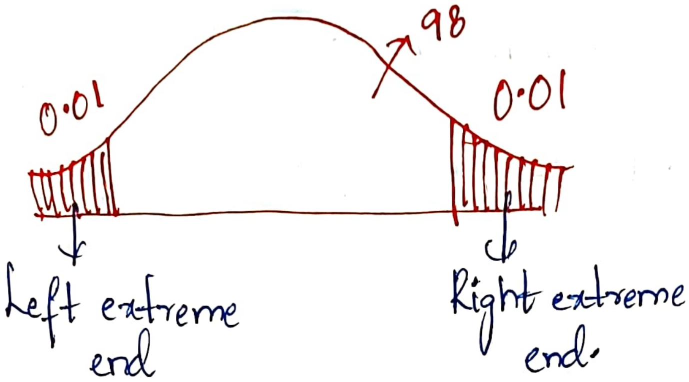
Given confidence region → 98

Critical region → 100 − 98

= 2

Two-tailed → $Z_{\alpha/2}$ ⇒ $\alpha = 0.02$

$$\frac{\alpha}{2} = \frac{0.02}{2} = 0.01.$$



It is sufficient that we find the z-score for one extreme end. So we redefine the confidence area as follows:

100 − 0.01

= 99.99

fraction part → $\boxed{0.99}$

Search it in the table.

⇒

	0.01	0.02	$\boxed{0.03}$	0.04	0.05	0.06	0.07
0.0							
0.1							
0.2							
⋮							
$\boxed{2.3}$			0.9901				

$$\Rightarrow p - z_{\alpha/2}\sqrt{pq/n} < P < p + z_{\alpha/2}\sqrt{pq/n}$$

$$9 < P < 16$$

Suppose we are asked for 96 % Confidence interval

Confidence region → 96

Critical " → 100 - 96

= 4

4 % LOS ⇒ $Z_{\alpha/2} = \frac{0.04}{2} = 0.02$

New confidence region → 100 - 0.02

= 99.98

fraction par → 0.98

Search in table

$Z_{\alpha/2} = 2.05.$

i.e., $9 < P < 16$

There will 9 to 16 bad apples out of 100 in the consignment.

6) Experience had shown that 20% of a manufactured product is off top quality. In one day's production of 400 articles, 50 are of top quality. Show that the either production of the day, taken was not a representative sample or the hypothesis of 20% was wrong.

Soln:

$H_0: P = 0.2$

$H_1: P \neq 0.2$

$$z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

$x = 50 \quad n = 400$

$$p = \frac{x}{n} = \frac{50}{400} = 0.125$$

$P = 0.2$

$Q = 1 - 0.2 = 0.8$

$$z = \frac{0.125 - 0.2}{\sqrt{0.2 \times 0.8 / 400}}$$

$= -3.75$

Table value of z at 5% LOS for two tailed test 1.96

18.01.2021

7) In a sample of 1000 people in karnataka 540 are
** rice eaters and rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance.

Sol: $H_0 : P = 0.5$ → as rate of wheat eaters = rice eaters.

$H_1 : P \neq 0.5$

$x = 540$

$n = 1000$

$P = 0.5$

$Q = 1 - P$
$= 1 - 0.5$
$= 0.5$

$$p = \frac{x}{n} = \frac{540}{1000} = 0.54$$

$$z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

$$= \frac{0.54 - 0.5}{\sqrt{(0.5)(0.5)/1000}} = 2.5298$$

Table value at 1% level of signifince, for a two tailed test is 2.5298

∴ Calculate value < Table value

Accpect H_0.

Infere : Hence both rice and wheat are equally popular in k'nataka

8) In a big city 325 men out of 600 men, were found to be cola drinkers. Does this information support the conclusion that majority of men in this city are cola drinkers.

Solⁿ: $H_0 : P = 0.5$

$H_1 : P > 0.5$

$x = 325$ $\quad p = \frac{x}{n} = \frac{325}{600} = 0.5416$

$n = 600$

$P = 0.5$ $\quad Q = 1 - P$

$= 1 - 0.5 = 0.5$

$$z = \frac{0.5416 - 0.5}{\sqrt{0.5 \times 0.5 \cancel{} / 600}} = 2.03797 \simeq 2.0380$$

Total value of z at 5% level of significance is for a right tailed test is 1.65

Calculated value > Total value

Hence,

Rejected H_0.

Thus :,

Inference : Majority of men are cola drinkers.

∴ Inference : Majority of men are cola drinkers.

9) In a random sample of 160 workers exposed to a certain amount of radation, 24 experienced some ill effects, construct a 99% confidence interval for corrosponding true proportion ?

confidence interval - is always 2 tailed test

Sol: $n = 160$

$x = 24$

$p = \frac{x}{n} = \frac{24}{160} = 0.15$

$q = 1 - p = 1 - 0.15$

$= 0.85$

$\alpha = 1\% = 0.01$

$z_{\alpha/2}$ at 1% for a two tailed test is 2.58

99% confidence interval for true proportion is

$$= p - z_{\alpha/2}\sqrt{pq/n} < P < p + z_{\alpha/2}\sqrt{pq/n}.$$

$$= 0.15 - 2.58\sqrt{\frac{0.15 \times 0.85}{160}} < P < 0.15 + 2.58\sqrt{\frac{0.15 \times 0.85}{160}}$$

$$= 0.77169 < P < 0.2228$$

Test for difference of Proportions :

$H_0 : P_1 = P_2$	$H_0 : P_1 = P_2$	$H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$	$H_1 : P_1 > P_2$	$H_1 : P_1 < P_2$

1) $$\boxed{z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)}$$

where,

p_1 = I^{st} sample proportion

p_2 = II^{nd} sample proportion

$$\boxed{P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}}$$

2) $$\boxed{Z = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}} \sim N(0,1)$$

$Q = 1 - P$

n_1 = I^{st} sample size

n_2 = II^{nd} sample size

P_1 = I^{st} Population Proportion

P_2 = II^{nd} Population Proportion

$Q_1 = 1 - P_1$

$Q_2 = 1 - P_2$

3) Confidence interval for difference of proportion

$$\boxed{(p_1 - p_2) \pm Z_{\alpha/2} \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

1) Random samples of 400 men and 600 women were asked wheather they would like to have fly over near their residence.

200 men and 325 women, were in favour of the proposal, test the hypothesis that proportion of men and women of the proposal are same at 5.% level.

Soln :

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2$$

So,

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

$= n_1 = 400, \quad n_2 = 600$

$x_1 = 200 \quad X_2 = 325$

$$p_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.5$$

$$p_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.5416$$

$$P = \frac{x_1 n_1 + x_2 n_2}{n_1 + n_2}$$

$$= \frac{200 + 325}{400 + 600} = 0.525$$

$Q = 1 - P = 1 - 0.525$

$= 0.475$

$$z = \frac{0.5 - 0.5416}{\sqrt{0.525 \times 0.475 \left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$= -1.2905$

Tab value z at 5 % level of significance for a two tailed is 1.96

Calculated value < Total value

Accepect H_0.

2) A candidate for election made a speech in a city A but not in B. A sample of 500 voters from a city A showed that 59.6 % of voters were in favour of him, where as sample of 300 voters from city B, showed that 50 % of the voters favored him, Discuss wheather his speech would produce, any effect on voters in city A.

ans) $H_0: P_1 = P_2$

$H_1: P_1 > P_2$

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

$n_1 = 500 \quad n_2 = 300$

$$x_1 = \frac{59.6}{100} \times 500 = 298$$

$$x_2 = \frac{50}{100} \times 300 = 150$$

$$p_1 = \frac{x_1}{n_1} \Rightarrow \frac{298}{500} = 0.596$$

$$p_2 = \frac{x_2}{n_2} \Rightarrow \frac{150}{300} = 0.5$$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{298 + 150}{500 + 300} = 0.56$$

$$Q = 1 - P \Rightarrow 1 - 0.56 = 0.44$$

$$z = \frac{0.596 - 0.5}{\sqrt{0.56 \times 0.44 \left(\frac{1}{500} + \frac{1}{300}\right)}}$$

$$= 2.6482$$

Tab value of z at 5% LOS for a right tailed test is 1.65

Cal val > Tab val

Rejected H_0.

4) In two large populations there are 30% and 25%, respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900, respectively in two populations?

Sol: H_0 : The difference in population is likely to be hidden in samples.

H_1 : The difference in population is not likely to be hidden in samples.

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_1 Q_2}{n_2}}} \sim N(0,1)$$

$P_1 = \frac{30}{100} = 0.3$ $\quad$ $P_2 = \frac{25}{100} = 0.25$

$Q_1 = 1 - P_1$
$= 1 - 0.3$
$= 0.7$

$Q_2 = 1 - P_2$
$= 1 - 0.25$
$= 0.75$

$n_1 = 1200$ $n_2 = 900$

$$z = \frac{0.3 - 0.75}{\sqrt{\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900}}} = 2.5538.$$

Table value of z at 5% level of significance for a two tailed test is 1.96

2 tailed, Right tailed	Left tailed test
$\rightarrow$ cal val < Table value	cal val < Tab val
Accpet Ho.	Reject H_0
Else,	otherwise
cal val > Tab val	cal val > Tab val
Reject H_0.	Accpect H_0

Cal val > Tab val.

Hence Reject H_0.

Inference,

The defference in population proportion is not likely in hidden samples.

Test for single Mean 22|01|2021

Consider

$H_0 : \mu = 30$	$H_0 : \mu = 30$	$H_0 : \mu = 30$
$H_1 : \mu \neq 30$	$H_1 : \mu > 30$	$H_1 : \mu < 30$

1) $$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$\bar{x}$ - sample Mean

μ = population Mean

σ = Population standred deviation.

n = sample size

(we use s, the sample standred deviation if σ is, not given since the sample size is large)

→ $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$

★

) Confidence interval for true Mean

$$\bar{x} - z_{\alpha/2}\, \sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2}\, \sigma/\sqrt{n}$$

$$\bar{x} - z_{\alpha/2}\, s/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2}\, s/\sqrt{n}$$

) According to Norms, established for a mechanical aptitude test persons who are 18 years old have height of 73.2 with a standred deviation of 8.6. If randomey selected of that age averaged 76.7; test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at 0.01 level of significance:

Solⁿ: $H_0: \mu = 73.2$

$H_1: \mu > 73.2$

$\mu = 73.2$

$\sigma = 8.6$ $\quad \bar{x} = 76.7$

$n = 4$

Prob: (1) According to norms established for a mechanical aptitu-de test, persons who are 18 yrs old have an avg height of 73.2 with a S.D of 8.6. If 4 randomly selected persons of that age averaged 76.7, test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at 0.01 LOS.

Soln:

$H_0 : \mu = 73.2$

$H_1 : \mu > 73.2$

$\mu = 73.2$

$\sigma = 8.6$

$n = 4$

$\bar{x} = 76.7$

$n \geq 30$ large Sample

$n < 30$ small "

So for large sample, if popn SD σ is not given, it can be replaced with sample SD, s.

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$= \frac{76.7 - 73.2}{8.6 / \sqrt{4}} = 0.81395$$

$$\simeq 0.8140.$$

Tab Val of z at 1% LOS for a right tailed test is 2.33

Cal Val < Tab Val

Accept H_0

Inference: $\mu = 73.2$ is true.

(2) A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a popⁿ with mean weight 56 kgs and S.D 25 kgs.

$H_0: \mu = 56$ or the sample is regarded to be drawn from a popⁿ with mean 56.

$H_1: \mu \neq 56$

(or)

The sample is not regarded to be drawn from a popⁿ with mean 56.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$\bar{x} = 70$, $\mu = 56$

$\sigma = 25$, $n = 64$

$$z = \frac{70-56}{25/\sqrt{64}} = 4.48$$

Tab Val of z at 5% LOS for a two tailed test is 1.96

Cal Val > Tab Val

Reject H_0

Inference: $\mu \neq 56$ is true

(or)

The sample is not regarded to be drawn from a popⁿ with mean 56.

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$\bar{x} = 70 \quad \mu = 56 \quad \sigma = 25 \quad n = 64$

$$Z = \frac{70 - 56}{25/\sqrt{64}} = 4.48$$

Table value of Z at 5% level of significance for a two tailed test is 1.96.

cal value > Table value

Rejected H_0

Inference,

$\mu \neq 56$, is true or the sample is not regarded to be drawn from a population with mean of 56 kgs.

3) A insurance agent has claimed that average age of policy holders who issue to him is the less than, the average for all agents which is 30.5 yrs. A random sample, of 100 policy holders who had issued through him, gave the following age distribution

Age	16-20	21-25	26-30	31-35	36-40
No. of Persons (f)	12	22	20	30	16

assume a particular, set of values $x_1, x_2,$

Calculate the arethmatic Mean and standred deviation of this distribution and use these values to test the claim at 5% loss of significance.

Sol^n: $H_0 : \mu = 30.5$

$H_1 : \mu < 30.5$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

To find mean & standred deviation of the sample

$$\text{Mean} = A + \frac{\Sigma f_i d_i}{N} \times h$$

$$\text{Standred deviation} = \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2} \times h$$

A - assumed mean

$d_i =$

C.I	f	m_i	$d_i = \frac{m_i - A}{h}$	$f_i d_i$	d_i^2	$f_i d_i^2$
15.5 - 20.5	12	18	$\frac{18-28}{5} = -2$	-24	4	48
20.5 - 25.5	22	23	-1	-22	1	22
25.5 - 30.5	20	28 → A	0	0	0	0
30.5 - 35.5	30	33	1	30	1	30
35.5 - 40.5	16	38	2	32	4	64
	N = 100			16		164

→ Middle term of should be taken

A - should be taken from m_i only, dont take from f

h = width of class interval

= upper limit - lower limit = 20.5 - 15.5 = 5

$\Sigma f_i = N = 100$ $\quad$ m_i – Mid point of class interval (C.I)

$\Sigma f_i d_i = 16$ $\quad$ $= \dfrac{\text{upper limit + lower limit}}{2}$

$\Sigma f_i d_i^2 = 164$

$$\text{Mean} = 28 + \frac{16}{100} \times 8 = 28.8$$

Standred deviation

$$= \sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2} \times 5$$

$$= \sqrt{1.64 - (0.16)^2} \times 5$$

$$= 6.3530$$

$\bar{x} = 28.8$ $\quad$ $s = 6.3530$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

$$= \frac{28.8 - 30.5}{6.3530/\sqrt{100}} = -2.67591$$

The value of z at 5% level of signifance for a left tailed test is -1.65

Cal val < Tab val

Reject H_0

Inference

$\mu < 30.5$ is true or the insurance agent claim is true.

4) The Mean & SD of a population are 11795 and 14054 respectly. If $n = 50$, find 95% confidence interval for mean.
***

Soln: Given,

$$\mu = 11795$$

$$\sigma = 14054 \quad n = 50$$

$Z_{\alpha/2}$ at 5% level of signficance or 95% confidence $= 1.96$

Confidence interval for true mean is

$$= \mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 11795 - 1.96 \times \frac{14054}{\sqrt{50}} < \mu < 11795 + 1.96 \times \frac{14054}{\sqrt{50}}$$

$$= 7899.4 < \mu < 15690.57$$

5) A random sample size 81 was taken whose varience is 20.25 and mean 32. Construct 98% confidence interval.

Soln: $n = 81$

Sample varience $= s^2 = 20.25$

sample SD $= S = \sqrt{20.25}$

$$\bar{x} = 32$$

$Z_{\alpha/2}$ at 98% confidence at 2% loss of significance $= 2.33$

confidence unterval for true mean μ.

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

$$32 - 2.33 \times \frac{4.5}{\sqrt{81}} < 32 + 2.33 \times \frac{4.5}{\sqrt{81}}$$

$$30.835 < \mu < 33.165$$

Test for difference of Means

25.01.2021

sample 1

$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 > \mu_2$	$H_1: \mu_1 < \mu_2$

1) $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1).$$

where

- $\bar{x}_1$ - first sample Mean
- $\bar{x}_2$ - second sample Mean

s_1 - sample standred deviation

s_2 - second standred deviation

n_1 - first sample size

n_2 - second sample size.

2) $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

where

x_1, x_2 — sample Mean

n_1, n_2 - Sample size.

σ - Population standred deviation

i.e., σ^2 = varience

3) Confidence interval formula

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

1) A survey of buying habits of $\overset{n_1}{400}$ women are choosen at random. shoppers located in certain section of a city. Their average weekly food expidenture is Rs. $\overset{\bar{x}_1}{250}$ with the standred deviation Rs $\overset{s_1}{40}$. For $\overset{n_2}{400}$ women shoppers choosen at random in a supermarket B, in another section of city of average food expidenture is Rs $200^{\bar{x}_2}$ with a standred deviation of Rs 55. Test at 10%. level of signifance wheather the average weekly food expenditure of two populations of shoppers are equal.

$H_0 = \mu_1 = \mu_2$ $\quad H_1 : \mu_1 \neq \mu_2$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

Here $\bar{x}_1 = 250$ $\quad \bar{x}_2 = 220$

$s_1 = 40$: $s_2 = 55$

$n_1 = 400$ $\quad n_2 = 400$

$$z = \frac{250 - 220}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{400}}}$$

$= 8.8226$

∴ Tab. value z at 10%, level of signif. for a two tailed test is 1.65,

cal val > Tab val.

Inference : The average weekly expenditure is not same in two population of shoppers.

2) Samples of students were drawn from two universities & from their weights in kgs, mean and deviation are calculated as shown below.

Test : $H_0 : \mu_1 \neq \mu_2$ against $H_1 : \mu_1 > \mu_2$

	Mean	SD	Size of Sample
University A	55	10	400
University B	57	15	100

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

$\bar{x}_1 = 55 \quad x_2 = 57 \quad n_1 = 400 \quad n_2 = 100$

$s_1 = 10 \quad s_2 = 45$

$$z = \frac{55 - 57}{\sqrt{\frac{(10)^2}{400} + \frac{(15)^2}{100}}} = -1.2649$$

Tab val of z at 5%.

Inference $\mu_1 = \mu_2$ is true

There is no significant difference between the two universities w.r.t the weight of students.

3) The mean Consumption of food ~~[illegible]~~ grains among 400 Samples middle class Consumers are 380 gms per day per person with a SD of 120 gm. A similar Sample survey of 600 working class Consumers

gave a mean of 410 gms with SD of 80 gms.
Can we say that working consume more food than the middle class consumers.

Sol: $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

$\mu_1 \rightarrow$ Middle class consumes

$n_1 = 400$, $x_1 = 380$

$s_1 = 120$

$\mu_2 \rightarrow$

$n_2 = 600$ $x_2 = 410$

$s_2 = 80$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

$$= \frac{380 - 410}{\sqrt{\frac{(120)^2}{400} + \frac{(80)^2}{600}}} = -4.3916$$

Tab value at 5% level of significance for a left tailed test is -1.65

cal val < Tab val

Rejected H_0.

inference : working class consume more food grains than middle class.

The means of two large samples of sizes 1000 & 2000 ~~sizes~~ are 67.5 inches, 68 inches respectively. Can be samples be regarded, as drawn from the ~~same~~ same population of standard deviation 2.5 inches?

Sol: $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$Z = \frac{\bar{x}_1 - \bar{x}_1}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim N(0,1)$$

$\bar{x}_1 = 67.5 \quad \bar{x}_2 = 68$

$n_1 = 1000 \quad n_2 = 2000$

$\sigma = 2.5$

$$Z = \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)}{2000}}}$$

$= -5.1640$

Since, it is a two-tailed test, $|Z| = 5.1640$

Tab value z at 5% level of significance for a two tailed test is 1.96

$\therefore$ cal val > Tab val.

Reject H_0

samples are not drawn from sample population with S.D 2.5

SMALL SAMPLE TEST

χ^2 (chi-square) : The square of a standred normal varient is a chi-square.

pdf of χ^2 :

Γ | $\Gamma \rightarrow$ Gamma function

$$f(\chi^2) = \frac{1}{2^{n/2}\,\Gamma\frac{n}{2}} e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1}$$

$; \; 0 < \chi^2 < \infty.$