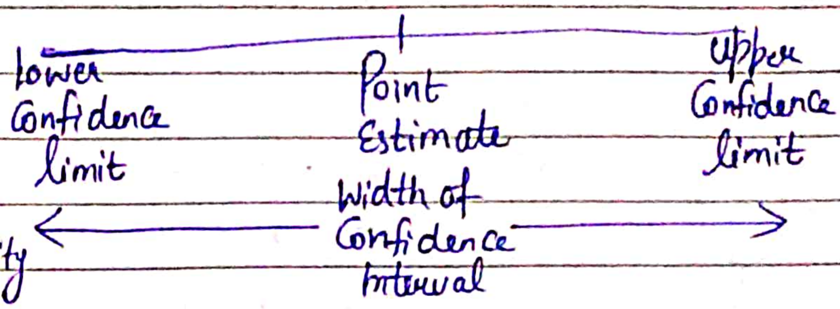
→ A Confidence Interval provides additional Information about the variability of the estimate.




Ex For a pop$^n$ with known variance of 185, a sample of 64 Individuals leads to 217 as an estimate of the mean.

(a) Find the std. error of the mean.

(b) Establish an Interval estimate that should Include the pop$^n$ mean 68.3% of the time.

Sol$^n$

$\sigma^2 = 185 \Rightarrow \sigma = \sqrt{185} = 13.60$

$n = 64; \quad \bar{x} = 217.$

$(\bar{x} \pm Z_{\alpha/2}\, \sigma_{\bar{x}})$

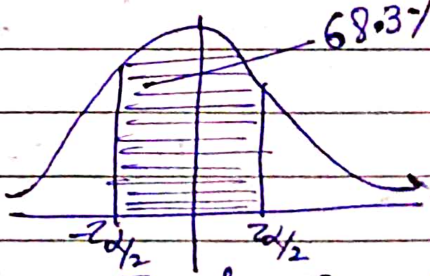
(a) $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{13.64}{\sqrt{64}} = 1.70$

$68.3\% = \dfrac{0.683}{2} = 0.3415$

(b) $\bar{x} \pm 1 . \sigma_{\bar{x}}$

$= \cancel{\bar{x}} + 217 \pm 1.70$

$= (215.3, 218.7)$




Z value Corr. to 0.3415 $= 1$

So $Z_{\alpha/2} = 1$

→ In the same Que, find the lower & upper Confidence limits for (a) 54% (b) 75% (c) 94% (d) 98%.

(a) $\dfrac{54}{100} = 0.54$ ; $\dfrac{0.54}{2} = 0.27 = \alpha/2$

$Z_{\alpha/2} = 0.74$

Date ______
Page ______

So Confidence Interval is $\bar{x} \pm z_{\alpha/2}\,\sigma_{\bar{x}}$

$$= 217 \pm 0.74\,(1.70)$$

Similary for 75% ; $\bar{x} \pm 1.15\,\sigma_{\bar{x}}$
for 94% ; $\bar{x} \pm 1.88\,\sigma_{\bar{x}}$
for 98%; $\bar{x} \pm 2.33\,\sigma_{\bar{x}}$.

* Thus ; Confidence Interval measures how much uncertainty is associated with a point estimate of the popn parameter?
~~for~~ Or Confidence Interval is the range within which popn parameter will fall?.

* Level of Confidence / Confidence Level is the Prob. that Popn parameter will ~~b~~ fall within this range
In General, 95% level of Confidence is assumed in real world applications. and other Could be 90%, 99%

General formula for all Confidence Intervals is

Point Estimate $\pm$ (Critical value) (Std. Error)

Where critical value is a table value based on the sampling distn of the Point Estimate and the desired Confidence level.

* Confidence Level — $(1-\alpha)$
If Confidence level is 0.95
i.e. $1-\alpha = 0.95$
$\Rightarrow \alpha = 0.05$.

→ In Short, we prepare a Confidence Interval for a Single sample of size ($n$) — say.

Based on one sample we selected, we can be 95% Confident (level of Confidence) that our Interval will ~~Confidence~~ Contain the popⁿ parameter.

| Confidence level | Confidence Coeff $(1-\alpha)$ | $Z_{\alpha/2}$ value |
|---|---|---|
| 90% | 0.90 | 1.645 |
| 95% | 0.95 | 1.96 |
| 99% | 0.99 | 3.27 |

# Hypotheses testing

→ Hypotheses are tentative Explanations of a principal operating in nature.
i.e. A statement of what the researcher believes will be the outcome of an experiment or a study.

→ Statistical hypotheses Consist of two parts

- Null hypothesis
- Alternative hypotheses.

→ Null hypothesis States that the null Condition exists.
i.e. there is nothing new happening, the old theory is still true, the old standard is Correct and the system is in Control. — Ho

→ Alternative hypothesis — States that the new theory is true, there are new standards, the system is out of Control and/or something is happening. — Ha

Eg. A Company's market share is 18%. Because of Increased marketing effort, Company officials believe the Company's market shares to be Increased by 18%.

Let p – Popⁿ proportion

then $H_o : p = 0.18$

$H_a : p > 0.18$

→ New Idea / New theory that Company officials want to prove is stated in the alternative hypothesis. ★

Date ______
Page ______

## What is hypothesis?

↓

A claim (assertion) about a popⁿ parameter.

**Popⁿ mean**

Eg. The mean monthly cell phone bill in this city = 420 Rs.

**Popⁿ Proportion**

Eg. The Proportion of adults in this city with cell phones = 0.68 or 68

## Null hypothesis – Ho

↓

- States the claim or assertion to be tested.

  E.g. $H_0 : \mu = 420$

↓

- Is always about a popⁿ parameter not about a sample statistic

  E.g. $H_0 : \mu = 420$ ~~$H_0 : \bar{X} = 420$~~

↓

- Begin with the assumption that the null hyp. ($H_0$) is true
- ~~Contains~~ Refers historical value. (old theories are true)
- Contains Always "=", "≥" or "≤" sign.
- May or may not be rejected.

## Alternative hypothesis ($H_a$)

↓

Is opposite of the null hypothesis

E.g. $H_a: \mu \neq 420$. or $H_a: \mu > 420$
or $H_a: \mu < 420$

↓

It challenges the old theories / Existing system

↓

May or may not be proven

↓

Generally, the hypothesis that the researcher is trying to prove.

★ **Statistical Inference**

- **Estimation** → Estimate the Pop$^n$ parameter from the sample.
- **Hypothesis Testing** → We assume about Pop$^n$ parameter and then we try to prove or dis prove it.

**Hypothesis Testing Process:** → – Claim: Pop$^n$ mean is 100.

$H_0: \mu = 100$ ; $H_a: \mu \neq 100$.

– Sample the pop$^n$ and find the sample mean $\bar{x}$.

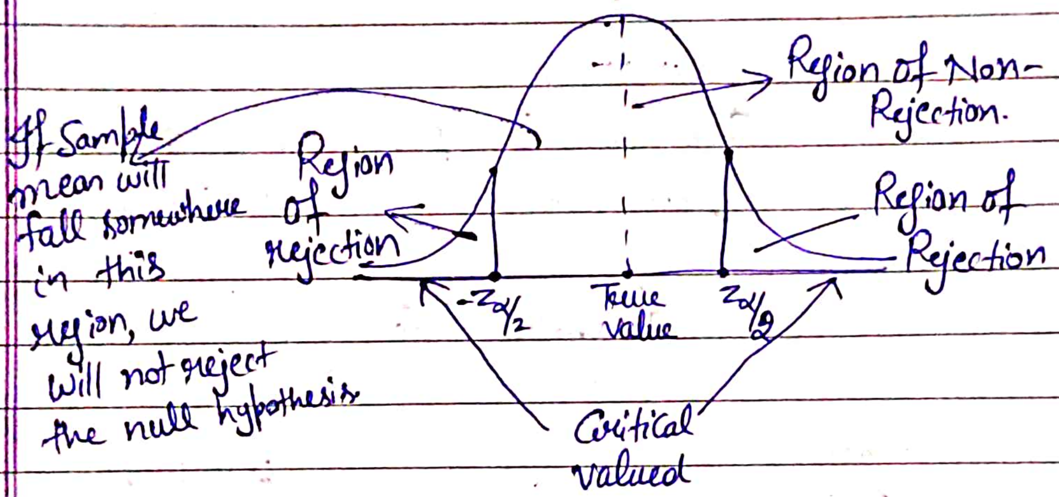
– If let $\bar{x} = 50$ then we reject the null hypothesis $H_0$.

[ ∵ $\bar{x}$ is significantly lower than the claimed Pop$^n$ mean $= 100$. ]

i.e. getting a sample mean 50 if $pop^n$ mean is 100 is so unlikely that we reject the null hypothesis $H_0$ ~~(that $pop^n$ mean)~~ and conclude that $pop^n$ mean must not be 100.

→ If the sample mean is close to the assumed $pop^n$ mean then null hypothesis is not rejected.

- And if the sample mean is far from the $pop^n$ mean then the null hypothesis is rejected.




## Errors In hypothesis Testing:→

As sample statistics are used to reach conclusion about $pop^n$ parameters, hence it is possible to make an incorrect decision about the null hypothesis.

- Errors
  - Type I
  - Type II

- Type I Error – When the null hypothesis is true and the decision maker decides to reject it.

→ The Prob. of Committing a type I error is Called ($\alpha$) or level of significance.

where $\alpha$ = area under the Curve that lies in the rejection region.

- Considered a Serious type of error.

<u>Type II error</u> When the null hypothesis is false and the decision maker fails to reject it.

- The Prob. of type II error is $\beta$.

| Decision | Actual Situation: $H_0$ True | Actual Situation: $H_0$ False |
|---|---|---|
| Do not Reject $H_0$ | No Error<br>Prob $(1-\alpha)$ | Type II Error<br>Prob $\beta$ |
| Reject $H_0$ | Type I Error<br>Prob $= (\alpha)$ | No error<br>Prob $(1-\beta)$ |

<u>Power of Test</u> – Prob of rejecting the null hypothesis $H_0$ when the null hypothesis is false. i.e. $(1-\beta)$.

- Confidence Coeff $(1-\alpha)$ – not rejecting $H_0$ when it is true.
- Confidence level $(1-\alpha) * 100\%$.

# Hypothesis test for the mean ($\mu$) :→

$\sigma$ Known (Z-Test) | $\sigma$ Unknown (t-Test)

Test Statistic is

$$Z = \frac{\bar{X} - \mu}{(\sigma/\sqrt{n})}$$

Step I Write $H_0$ and $H_a$.

Step II Choose level of significance $\alpha$.

Step (III) Convert $\bar{X}$ — to Z (test Statistic)

(IV) Determine the critical Z values for a specified level of Significance $\alpha$ from a table.

(V) If the test statistic falls in the rejection region, reject $H_0$, otherwise do not reject $H_0$.

**Que** Test the claim that mean no. of TV Sets in Indian homes = 3.

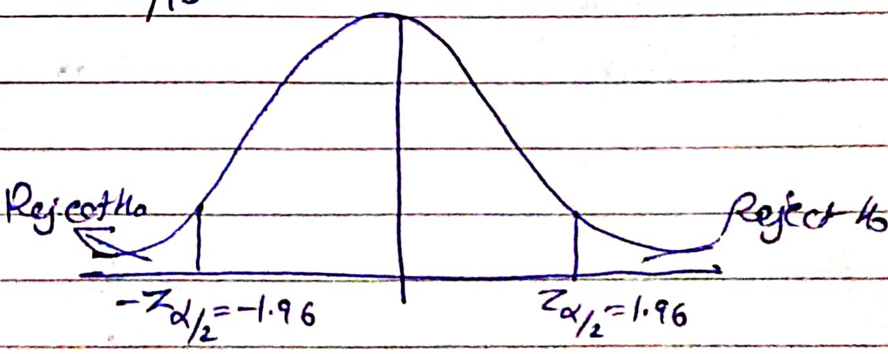
$\sigma = 0.8$; $\alpha = 0.05$; $n = 100$, $\bar{X} = 2.84$

**Sol$^n$** $H_0: \mu = 3$ ; $H_a: \mu \neq 3$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{2.84 - 3}{0.8/10} = -2.0$$

$\alpha = 0.05$

$\Rightarrow 1 - \alpha = 0.95$




Reject $H_0$ If

$Z < -1.96$ or

$Z > 1.96$

and $Z = -2.0 < -1.96$ (falls in Rejection Region)

$\Rightarrow$ Reject $H_0$.

**Que** Axel Strength is 80000 pounds per Square Inch

So $H_0: \mu = 80000$

$\sigma = 4000; \; n = 100, \; \bar{X} = 79600;$

$\alpha = 0.05$

**Sol^n**

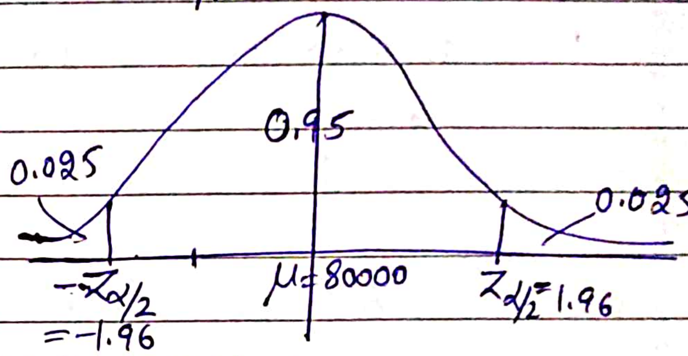
$H_0: \mu = 80000$

$H_a: \mu \neq 80000$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{79600 - 80000}{4000/10} = -1$$

$\alpha = 0.05$

$\Rightarrow 1 - \alpha = 0.95$




$\Rightarrow$ Z Statistics falls within non-Rejection region

$\Rightarrow$ Do not reject $H_0$.

**Que** Drug dose of 100cc has to be given to the patient. Excess dose is not harmful but Insufficient dose will not produce desired result.

$\sigma = 2, \; n = 50, \; \bar{X} = 99.75; \; \alpha = 0.10.$

$H_0: \mu = 100$

**Sol^n**

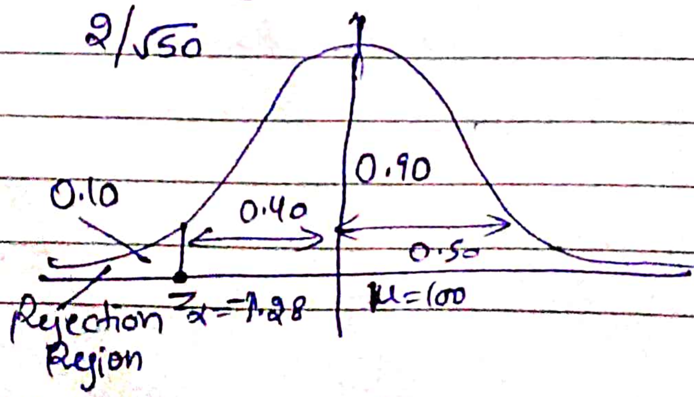
$H_0: \mu = 100$

$H_a: \mu < 100$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{99.75 - 100}{2/\sqrt{50}} = -0.88$$

(Look at Z table where area is 0.40)

$Z_\alpha = -1.28$



$Z_{value} = -0.88$ falls within non-rejection region
$\Rightarrow$ Do not reject null hypothesis $H_0$.

→ **when $\sigma$ is unknown:→** If $\sigma$ is unknown then we use the sample std. deviation S.

- We use t-distⁿ instead of z-distⁿ to test the null hypothesis about the mean.

Main assumption — The popⁿ from which samples are drawn follows a normal distⁿ.

- Remaining all steps and conclusions will be same.

The test statistic is

$$t_{value} = \frac{\bar{X} - \mu}{S/\sqrt{n-1}}$$

Que:→ The avg cost of a hotel room in Delhi is Rs. 168 per night. To check if it is true, a random sample of 25 hotels is taken and resulted in $\bar{X} = 172.50$ Rs. and $S = 15.40$ Rs. Test the appr. hypotheses at $\alpha = 0.05$.

Solⁿ $H_0: \mu = 168$ ; $H_a: \mu \neq 168$

$n = 25$; $\bar{X} = 172.50$ ; $S = 15.40$ ; $\alpha = 0.05$

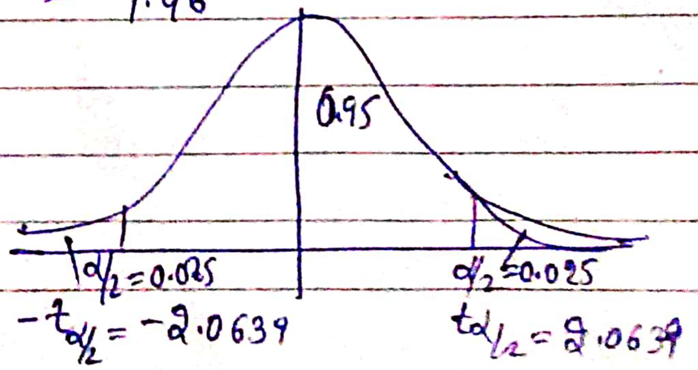
$$t_{value} = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{172.50 - 168}{15.40/\sqrt{25}}$$

$$= 1.46$$

With $(n-1)$ degrees of freedom

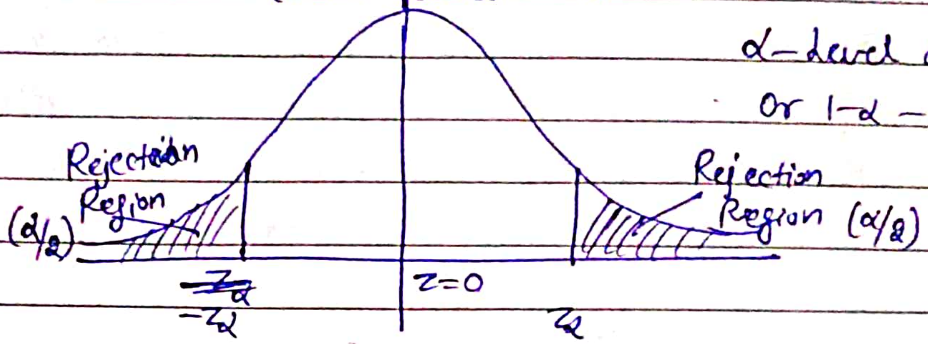
$t_{24, 0.025} = 2.0639$

$t_{value} = 1.46$ falls within



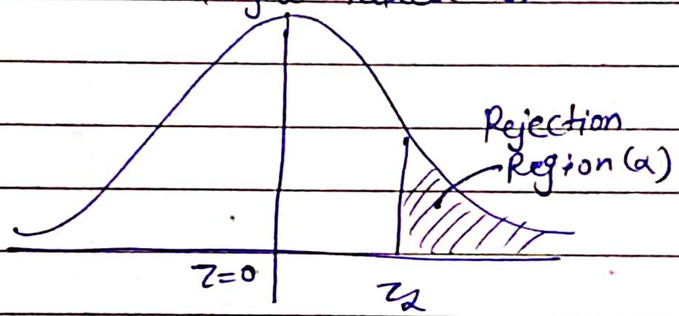
non-rejection region.
⇒ do not reject $H_0$.

Two Tailed Test

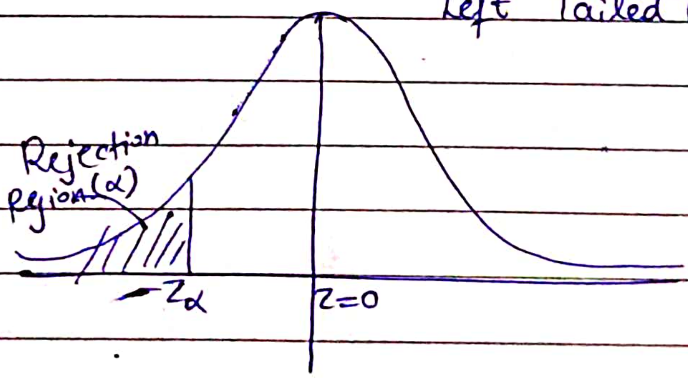



$\alpha$ – level of Significance
or $1-\alpha$ – level of Confidence.

Right tailed Test




Left Tailed Test




| | Level of Significance ($\alpha$) | | |
|---|---|---|---|
| Critical value ($z_\alpha$) | 1% | 5% | 10% |
| Two tailed test | $\lvert z_\alpha \rvert = 2.58$ | $\lvert z_\alpha \rvert = 1.96$ | $\lvert z_\alpha \rvert = 1.645$ |
| Right tailed test | $z_\alpha = 2.33$ | $z_\alpha = 1.645$ | $z_\alpha = 1.28$ |
| Left tailed test | $z_\alpha = -2.33$ | $z_\alpha = -1.645$ | $z_\alpha = -1.28$ |

**Que** A Sample of 900 members has a mean 3.4 Cms and S.d. 2.61 Cms. Is the sample from a large popⁿ of mean 3.25 Cms and S.d 2.61 Cms?

→ If the popⁿ is normal and mean is unknowns, find the 95% and 98% limits of true mean.

**Solⁿ**

$H_0: \mu = 3.25$ cm ~~and σ~~

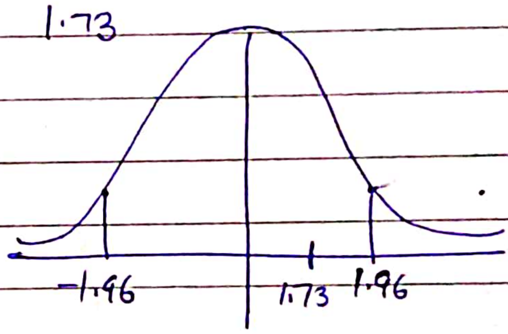
$H_1: \mu \neq 3.25$ Cm (Two tailed)

$$Z_{\text{Statistic}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.25}{2.61/\sqrt{900}}$$

$$= 1.73$$

$|Z| < 1.96$

⇒ We don't reject $H_0$ at 5% Level of Significance.




95% limits for popⁿ mean are

$$\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right) = 3.40 \pm 1.96\left(\frac{2.61}{\sqrt{900}}\right)$$

$$= (3.5705, 3.2295)$$

98% limits for popⁿ mean are

$$\bar{x} \pm 2.33\left(\frac{\sigma}{\sqrt{n}}\right) = 3.40 \pm 2.33\left(\frac{2.61}{\sqrt{900}}\right)$$

$$= (3.6027, 3.1973)$$

**Que** A machinist is making engine parts with axle diameters of 0.700 Inch. A random sample of 10 parts shows a mean diameter of 0.742 Inch with a std. dev. of ~~0.400~~ 0.040 Inch. Compute the statistic.

**Sol$^n$**

$H_0: \mu = 0.700$

$H_1: \mu \neq 0.700$

$\bar{x} = 0.742,\ n = 10,$

$S = 0.040$

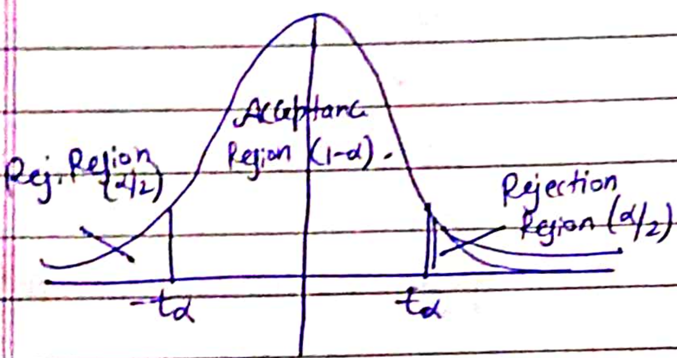
Test statistic is

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}} = \frac{0.742 - 0.700}{0.040/\sqrt{9}}$$

$$= 3.15 \sim t_{n-1,\,0.05}$$

$$= t_{9,\,0.05} = \cancel{3.15}$$

$$= 2.26$$




$3.15 > t$ value

$\Rightarrow$ Ho is rejected.

**Que** $\mu = 146.3$ — Sales.

$n = 22;\ \bar{x} = 153.7;\ s = 17.2;\ \alpha = 0.05$

**Sol$^n$**

$H_0: \mu = 146.3$

$H_1: \mu > 146.3$ (Right tail)

$$t_{statistic} = \frac{\bar{x} - \mu}{S/\sqrt{n-1}} = \frac{153.7 - 146.3}{17.2/\sqrt{21}}$$

$$= 2.03 \sim t_{0.05,\,21}$$

$$= t_{21}(0.10)$$

$$= 1.72$$

$\because$ tstat $>$ tvalue

$\Rightarrow$ reject Ho.

Date ____
Page ____

**Note** The significant values of $t$ at level of significance $\alpha$ for a single tailed test can be obtained from those of two tailed test by looking the values at level of significance $2\alpha$.

E.g. $t_8(0.05)$ for single tail test =
$t_8(0.10)$ for two-tail test $= 0.10$

$t_{15}(0.01)$ for single tail test =
$t_{15}(0.02)$ for two tail test $= 2.60$.

**Que** A random sample of 10 boys had the following IQ's.
70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

Do these data support the assumption of popⁿ mean I.Q. of 100? Find the reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

**Solⁿ**

$H_0: \mu = 100$

$H_1: \mu \neq 100$

$\bar{x} = 97.2$

$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = 183.36$

Calculate from given data.

$t_{statistic} = \frac{97.2 - 100}{\sqrt{183.36/9}} = 0.62 \sim t_{0.05,9}$

$t_{value} = 2.262$

$\because t_{stat} < t_{value}$

$\Rightarrow H_0$ is not rejected at 0.05 or 5% level of significance.

95% Confidence limits ~~are~~ are

$\bar{x} \pm t_{0.05} \cdot S/\sqrt{n} = 97.2 \pm 2.26(4.514)$

$= 107.41$ & $86.99$

$\Rightarrow$ Confidence Interval is $[86.99, 107.41]$

Testing of Proportion:→ If X is the no. of Individuals (units) possessing the give attribute (termed as a Success), with Constant Prob "P" of success, in a random sample of size $n$ from a large popⁿ then the Prob. of $x$ Successes in $n$ trials is given by the binomial distⁿ

$$P(X=x) = n_{C_x} P^x . Q^{n-x} ; x = 0,1,2 \dots n$$

where $P, Q \geq 0$ and $P+Q=1$.

If $n \to$ large, then binomial distⁿ tends to a normal distⁿ i.e. for large $n$, $X \sim N(nP, nPQ)$

$$Z_{stat} = \frac{X - E(x)}{\sqrt{Var\, x}} = \frac{X - nP}{\sqrt{nPQ}} \sim N(0,1).$$

$p$ — observed Sample proportion of the Success $= \frac{X}{n}$.

$P$ — Popⁿ proportion.

$$E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{nP}{n} = P$$

$$V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{nPQ}{n^2} = \frac{PQ}{n}$$

∴ for large sample, the standard normal variate Corresponding to the statistics "p" is

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

## Two-tailed test

Step I — Define $p$ — Sample Proportion

II — Define $H_0$ and $H_1$

III — Compute the value of $z = \frac{p-P}{\sqrt{PQ/n}}$

IV — If $|z| > 3$ then $H_0$ is rejected

If $|z| \leq 3$ then $H_0$ may not be rejected.

→ at '$\alpha$' level of Significance.

Compute the value of $Z_\alpha$

If $|z| > Z_\alpha$ then $H_0$ is rejected

If $|z| \leq Z_\alpha$ then $H_0$ is not rejected

Similarly we can do it for left or right tail test.

**Que** In a sample of 1000 people in Delhi, 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?

**Sol$^n$:-**

$n = 1000$

$X$ = no. of rice eaters = 540

$p = \frac{540}{1000} = 0.54$ — Sample prop. of rice eaters.

$H_0$: both rice and wheat are equally popular

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

$H_1$: $P \neq \frac{1}{2}$ (two-tailed)

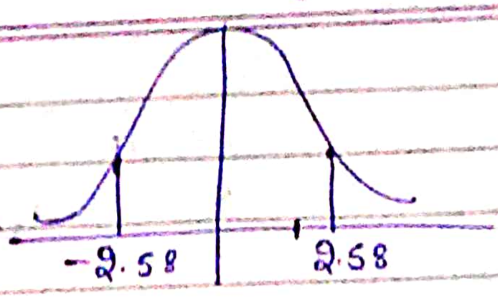
$$Z_{stat} = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54-0.5}{\sqrt{\frac{\frac{1}{2}\times\frac{1}{2}}{1000}}} = 2.532$$

$Z_{value}$ (at $\alpha = 0.01$) $= 2.58$

$Z_{stat} < Z_{value}$

$\therefore$ $H_0$ is not rejected.

ie rice & wheat are equally popular.




**Que** Twenty people were attacked by a disease and only 18 survived. Will You reject the hypothesis that the survival rate, if attacked by this disease, is 85% in favour of the hypothesis that it is more, at 5% level?

**Solⁿ** $n = 20$

$X$ = no. of survived people = 18

$p = \frac{X}{n} = \frac{18}{20} = 0.9$ — prop. of persons survived In the sample.

$H_0: P = 0.85$

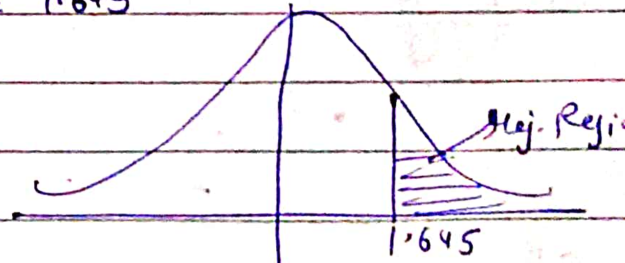
$H_1: P > 0.85$ (Right tail)

$$Z_{stat} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.9 - 0.85}{\sqrt{\frac{0.85 \times 0.15}{20}}} = 0.633$$

$Z_{value}$ (at $\alpha = 0.05$) = 1.645

$0.633 < 1.645$

$\Rightarrow$ $H_0$ is not rejected.




**Que** In a random sample of 400 persons from a large popⁿ, 120 are female. Can it be said that males and females are in the ratio 5:3 in the popⁿ at 1% level of significance?

Date ______
Page ______

**Sol^n**

$n = 400$

$x$ – no. of females $= 120$

$p = \frac{x}{n} = \frac{120}{400} = 0.30$

$H_0$: Males and females are in ratio 5:3

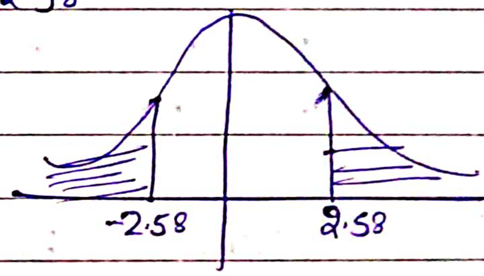
i.e. $P = \frac{3}{8} = 0.375$

$H_1$: $P \neq 0.375$ (two-tail)

$$Z_{stat} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.30 - 0.375}{\sqrt{\frac{0.375 \times 0.625}{400}}}$$

$$= -3.125$$

$Z_\alpha$ (at $\alpha = 0.01$) $= 2.58$




⇒ $Z_{stat}$ lies in the rejection region

⇒ reject $H_0$

i.e. Males and females are not in the ratio 5:3

**Que** In a big city, 325 out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?

**Sol^n**

$n = 600$ – (large)

$x$ – no. of smokers $= 325$

$p = \frac{325}{600} = 0.5417$ – prop of smokers

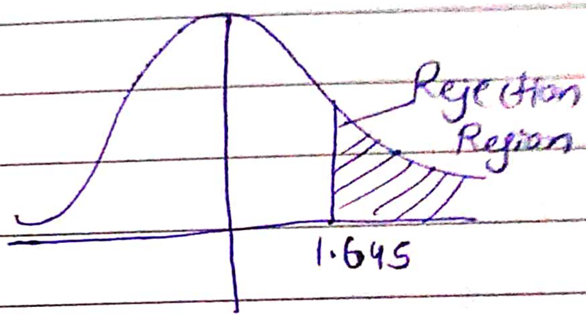
$H_0$: The no. of smokers and non-smokers are equal in the city.

i.e. $P$ = Pop$^n$ prop. of smokers in the city = 0.5

$H_1$ : $P > 0.5$ (right tailed)

$$Z_{stat} = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.5417-0.50}{\sqrt{\frac{0.5 \times 0.5}{600}}} = 2.04$$

$Z_\alpha$ (at 0.05) = 1.645




$Z_{stat} > Z_\alpha$

$\Rightarrow$ Ho is rejected

$\Rightarrow$ Majority of men are smokers.