Unit-4

Statistical Inference

→ Population
→ Sample → large – $n \geq 30$
→ small – $n < 30$

→ Parameters → μ, σ, P
→ Statistic → $\bar{x}$, s or S, p
→ Statistical Inference
→ Estimation → Testing of hypothesis
→ Point → Interval

$\hat{\mu} = \bar{x}$ $\quad 0.2 < P < 0.5$
$\hat{\sigma} = s$ or S
$\hat{P} = p$

→ **Testing of hypothesis :-**

Hypothesis :- Claim / Conjecture
→ Simple → composite

$\mu = 30$ $\quad \mu \geq 30$

→ **Testing of Single mean (Large sample) :-**

Step 1 :- Understanding the problem.
Identification of random variable.

Step 2 :- Stating the hypothesis

Null hypothesis H_0 : $\mu = \mu_0$
vs

Alternate hypothesis H_1 : $\mu > \mu_0$
$\mu < \mu_0$
$\mu \neq \mu_0$

Page No.
Date

Step 3 :- Test statistic

$$Z = \frac{\bar{x} - \mu_0}{S.E(\bar{x})}$$

Sampling distribution
→ drawing random samples from the population.

where $S.E(\bar{x})$ = Standard error$(\bar{x})$
= Standard deviation of $\bar{x}$
$= \sigma/\sqrt{n}$

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Step 4 :- Decision ~~Rate~~ Rule

α = level of significance = P(Type I error)
β = P(Type II error)

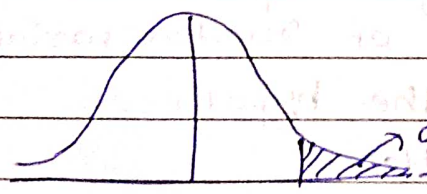
Eg :-

	Raining	Not raining
umbrella	✓	×
No umbrella	×	✓

$\alpha = 5\%$ or 1%

	Accept H_0	Reject H_0
True - H_0	✓	× → Type I error
False - H_0	× → Type II error	✓

case 1 : $H_1 : \mu > \mu_0$ (Right tail test)

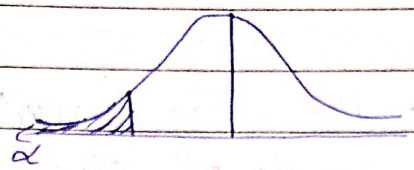


At $\alpha\%$, accept H_0 when

→ Critical / Rejection region

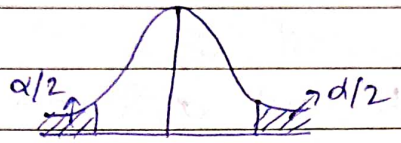
Accept H_0 when $Z_{calculated} < Z_\alpha$

Case 2 :- $H_1 : \mu < \mu_0$ (Left tail test)



Accept H_0 when $z > z_\alpha$

Case 3 :- $H_1 : \mu \neq \mu_0$



$|z| < z_{\frac{\alpha}{2}}$

Step 5 :- Numerical Computation

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$\bar{x} = \frac{\sum x}{n}$, $\hat{\sigma}$ (unknown) ← when popⁿ S.D is not known we replaced with its unbiased estimator

$$\hat{\sigma} = s = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2}$$

Conclusion, Accepting H_0 based on Step 4.

Page No.
Date

Test of hypothesis concerning single population μ for large sample.

Let μ and σ^2 be the mean & variance of a popn from which a random sample of size n is drawn.
Let $\bar{x}$ be the mean of sample for large sample from central limit theorem it follows that, sampling distn of $\bar{x}$ is approx. normally distributed with

$$\cancel{\mu = \bar{x} \text{ and } \sigma}$$

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

The test statistics is given by,

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

Problems:-

1) The length of life 'X' of certain computers is approx. normally distn with mean 800 hours & S.D 40 hours. If a random sample of 30 computers has an average life of 788 hours. Test whether there is a difference b/w avg. values i.e, test whether $\mu \neq 800$ at $\alpha = 5\%$ level of significance.

Step 1:- X - Life of a computer

$$X \sim N(\mu = 800, \sigma^2 = 40^2)$$

Step 2:- H_0 :- $\mu = 800$ (Avg. life of comp. is 800 hrs)

vs

H_1 :- $\mu \neq 800$ (Avg. life of comp. is not 800 hrs)

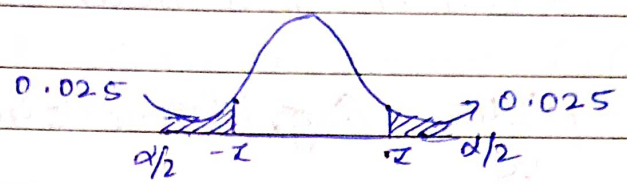
Page No.
Date

Step 3 :- Test Statistic.

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

Step 4 :- Decision rule.

At $\alpha = 5\%$, we accept H_0 when $|z| < Z_{\alpha/2}$



$$|z| < Z_{\alpha/2}$$

$$0.025 = 0.5 - P(0 < z < Z_{\alpha/2})$$

$$P(0 < z < Z_{\alpha/2}) = 0.475$$

$$\Rightarrow \boxed{Z_{\alpha/2} = 1.96} \rightarrow \text{critical value}$$

Step 5 :- Numerical Analysis/Computation

Given, $\mu = 800$, $\sigma = 40$, $n = 30$, $\bar{x} = 788$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{788 - 800}{40/\sqrt{30}}$$

$$z = -0.05477$$

$$z = -1.6431$$

Conclusion,

$$|z| = |-1.64| = 1.64$$

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

Since $|z| < Z_{\alpha/2}$, we accept H_0 i.e, average life of computers is 800 hours.

Page No.
Date

2) Mice with an average lifespan of 32 months will live upto 40 months when fed by a certain nutritious food. If 64 mice fed on this diet have an avg. lifespan of 38 months and S.D of 5.8 months. Is there any reason to believe that avg. lifespan is less than 40 months.

→ X - Lifespan of a mouse

→ H_0 :- $\mu \geq 40$ (Avg. lifespan ≥ 40 months)

VS

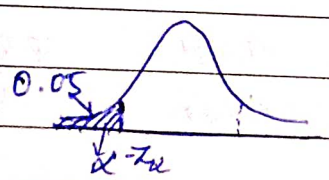
H_1 :- $\mu < 40$ (Avg. lifespan < 40 months)

→ Test statistic,

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

→ Decision rule,

At $\alpha = 5\%$, accept H_0 when $z > z_\alpha$



$0.05 = 0.5 - P(0 < z < z_\alpha')$

$P(0 < z < z_\alpha') = 0.45$

$\Rightarrow \boxed{z_\alpha' = 1.65}$

$\Rightarrow z_\alpha = -z_\alpha' = -1.65$.

→ Numerical computation,

$\mu = 40$, $n = 64$, $\bar{x} = 38$, $s = 5.8$

(Since σ is unknown, $\hat{\sigma} = s$)

$$z = \frac{38 - 40}{5.8/\sqrt{64}}$$

$z = -2.7586$

Page No.
Date

Conclusion,

$z = -2.76$ and

$z_\alpha = -1.65$

$\Rightarrow \quad z < z_\alpha$ or $z \not> z_\alpha$

Hence, We reject H_0

There is a reason to believe that avg. lifespan of mice is less than 40 months.

3) A pharmaceutical firm claims that the mean time for a drug to take effect is 24 ~~months~~ minutes. In a sample of 400 trials, the mean time is 26 ~~months~~ minutes with a S.D of 4 minutes. Test if the mean time for a drug to be effective is significant.

→ X- Time taken for a drug to be effective.

→ H_0 :- $\mu \geq 24$; drug is ineffective

vs

H_1 :- $\mu < 24$ drug is effective

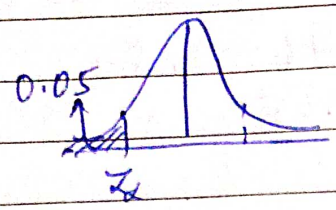
→ Test statistic,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

($\bar{x}$ → sample mean)

→ Decision rule,

At $\alpha = 5\%$, we accept H_0 when $z > z_\alpha$.



$Z > Z_\alpha$

$0.05 = 0.5 - P(0 < Z < Z'_\alpha)$

$P(0 < Z < Z'_\alpha) = 0.45$

$Z'_\alpha = 1.65$

$\Rightarrow \boxed{Z_\alpha = -1.65}$

→ Numerical Computation,

$\bar{x} = 26, \; n = 400, \; \sigma = 4, \; \mu = 24$

$$Z = \frac{26-24}{4/\sqrt{400}} = \frac{2}{4} \times 20 = 10$$

$\boxed{Z = 10}$

→ Conclusion,

$Z = 10$ and

$Z_\alpha = -1.65$

Since, $Z > Z_\alpha$

Hence, we can say that drug is ineffective.

4) In a health survey of school children it is found that the mean heamoglobin level of 55 boys is 10.2 with a SD of 2.1. Can we consider this group as taken from a population with a mean of 11?

→ X- Haemoglobin Level of Boys

Page No.
Date

→ H_0 :- $\mu = 11$ (Avg. Hb level of boys is 11)

vs

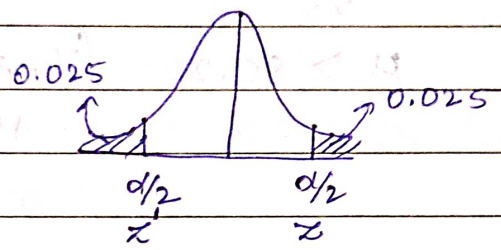
H_1 :- $\mu \neq 11$ (Avg. Hb level of boys $\neq$ 11)

→ Test statistic:-

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

→ Decision rule,

At $\alpha = 5\%$, we accept H_0 when $|z| < z_{\alpha/2}$



$|z| < z_{\alpha/2}$

$0.025 = 0.5 - P(0 < z < z_{\alpha/2})$

$\Rightarrow P(0 < z < z_{\alpha/2}) = 0.475$

$\Rightarrow \boxed{z_{\alpha/2} = 1.96}$

→ Numerical Analysis / Computation,

$N = 55$, $\mu = 11$, $\bar{x} = 10.2$, $s = 2.1$

$$z = \frac{10.2 - 11}{2.1/\sqrt{55}} = \frac{-0.8 \times \sqrt{55}}{2.1}$$

$$z = -2.829$$

→ Conclusion, $|z| = +2.829$ and $z_{\alpha/2} = 1.96$

$\Rightarrow z \nless z_{\alpha/2}$

Hence, we ~~accept~~ reject H_0 i.e., Average Haemoglobin level of Boys is ~~10.2~~ ~~11~~ not 11.

Page No.
Date

5) A new variety of potato grown in 250 plots gave rise to a mean yield of 82.7 quintals/hectare with a SD of 14.6 quintals/hectare. Is it reasonable to assert that the new variety is superior in yield to the standard variety with an established mean yield of 80.2 quintals/hectare?

→ X – Yield of potatoes

→ H_0 :- $\mu \leq 80.2$ (The new variety is not as superior)

vs

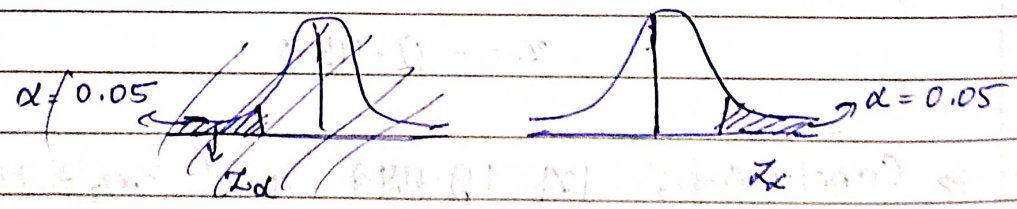
H_1 :- $\mu > 80.2$ (The new variety is superior than standard variety)

→ Testing statistic,

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

→ Decision rule,

At $\alpha = 5\%$, we accept H_0, when $z < z_\alpha$



$z < z_\alpha$

$0.5 - P(0 < z < z_\alpha) < 0.05$

$P(0 < z < z_\alpha) = 0.45$

$\Rightarrow z_\alpha = 1.64.$

Page No.
Date

→ Calculation,

$N = 250, \mu = 80.2, \bar{x} = 82.7, SD = 14.6$

$$Z = \frac{82.7 - 80.2}{14.6/\sqrt{250}}$$

$$Z = 2.7.$$

→ Conclusion,

$Z = 2.7$ and $Z_\alpha = 1.64$

$Z > Z_\alpha$

Hence, we reject H_0 i.e.,
The new variety is superior than standard variety.

6) A company claims that the mean thermal efficiency of diesel engines produced by them is 32.3. To test this claim a random sample of 40 engines were examined wch showed the mean thermal efficiency of 31.4 and SD of 1.6. Can the claim be accepted or not at 0.01 level of significance.

→ X – Thermal efficiency of diesel engines

→ H_0 :- $\mu = 32.3$ (Thermal efficiency of diesel engines is 32.3)
VS
H_1 :- $\mu \neq 32.3$ (— " — is not 32.3)

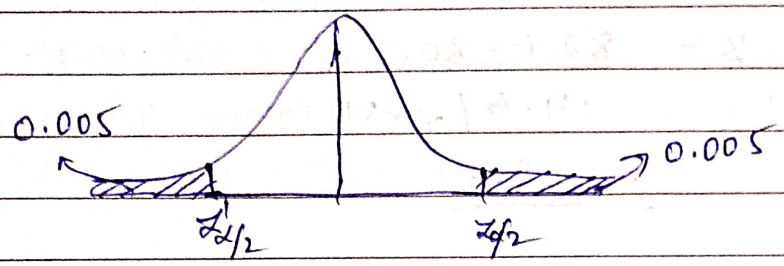
→ Testing statistic,

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

Page No.
Date

→ Decision rule,

At $\alpha = 1\%$, we accept H_0 when $|Z| < Z_{\alpha/2}$



$0.005 = 0.5 - P(0<Z<Z_{\alpha/2})$

$P(0<Z<Z_{\alpha/2}) = \cancel{0.45}$
$= 0.495$

$\Rightarrow Z_{\alpha/2} = 2.57.$

→ Calculation,

$N = 40, \mu = 32.3, \bar{x} = 31.4, s = 1.6$

$$Z = \frac{31.4 - 32.3}{1.6/\sqrt{40}}$$

$$\underline{Z = -3.55}$$

→ Conclusion,

$|Z| = +3.55$ and $Z_{\alpha/2} = 2.57$

$\Rightarrow Z \nless Z_{\alpha/2}$

Hence, we ~~accept~~ reject H_0 i.e.,
The thermal efficiency of diesel engines is not 32.3.

7) It has been previously recorded that the average depth of ocean at a particular region is 67.4 fathoms. Is there any reason to believe this at 0.01 level of significance (LOS) if the readings at 40 random locations in that particular region showed a mean of 69.3 with SD of 5.4 fathoms?

Page No.
Date

Test of significance concerning difference of mean (large sample).

* $X_1 \sim N(\mu_1, \sigma_1^2)$
  $X_2 \sim N(\mu_2, \sigma_2^2)$

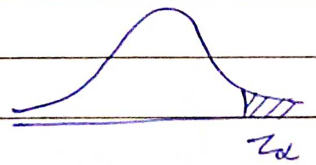
* $H_0 : \mu_1 = \mu_2$
  $H_1 : \mu_1 > \mu_2$
  $\mu_1 < \mu_2$
  $\mu_1 \neq \mu_2$

* Test statistic

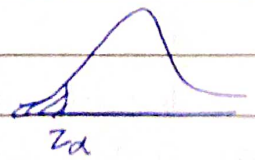
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

* Decision rule,

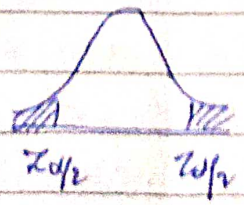
At $\alpha\%$, we accept H_0 when

→ 

$z < z_\alpha$

→ 

$z > z_\alpha$

→ 

$|z| < z_{\alpha/2}$

Page No.
Date

* Numerical computation,

$$\bar{x}_1 = \frac{\sum_i x_{1i}}{n_1} \qquad \bar{x}_2 = \frac{\sum_i x_{2i}}{n_2}$$

$$\sigma_1^2 = s_1^2 = \frac{1}{n_1} \sum_i (x_{1i} - \bar{x}_1)^2$$

$$\sigma_2^2 = s_2^2 = \frac{1}{n_2} \sum_i (x_{2i} - \bar{x}_2)^2$$

Problems :-

1) In a random sample of 100 tubelights produced by company A, the mean lifetime of tubelight is 1190 hours with S.D of 90 hours. Also in a random sample of ~~certi~~ 75 tubelights of company B, the mean lifetime is 1230 hrs with S.D of 120 hrs. Is there a difference b/w the mean lifetime of the 2 brands of tubelights at 0.01 LOS?

→ X_1 - lifetime of tubelight produced by company A
X_2 - —"— B

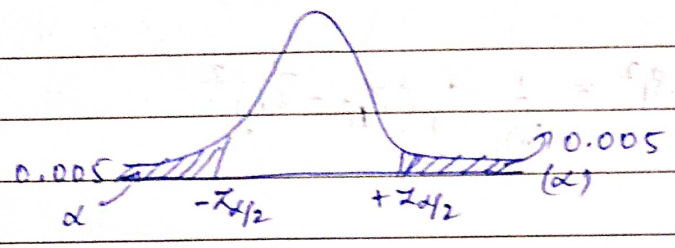
→ H_0 :- $\mu_1 = \mu_2$ (
vs
H_1 :- $\mu_1 \neq \mu_2$ (There is a difference b/w the mean lifetime of 2 brands)

→ Testing statistic,

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

→ Decision rule,

At $\alpha = 1\%$ LOS, we accept H_0 when, $|z| < z_{\alpha/2}$



$0.005 = 0.5 - P(0<z<z_{\alpha/2})$

$P(0<z<z_{\alpha/2}) = 0.495$

$\Rightarrow \boxed{z_{\alpha/2} = 2.57}$

→ Numerical computation,

$n_1 = 100,\ n_2 = 75,\ \bar{x}_1 = 1190,\ \bar{x}_2 = 1230,$

$s_1 = 90,\ s_2 = 120$

Replace, $\hat{\sigma}_1$ by s_1 & $\hat{\sigma}_2$ by s_2

$$z = \frac{(1190 - 1230) - \overset{(\mu_1 - \mu_2)}{\cancel{(1190 - 1230)}}}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}} \quad (\text{buoz } H_0 : \mu_1 - \mu_2)$$

$$z = \frac{-40}{\sqrt{273}} = \underline{-2.42}$$

→ Conclusion,

$z_{\alpha/2} = 2.57$

$|z| = 2.42$

Since $|z| < z_{\alpha/2}$

We accept H_0, i.e., there is no difference b/w the lifetimes of the 2 brands of tubelights.

2) To test the effect of a new pesticide on rice production a farmland was divided into 60 units of equal areas where all the portions having identical qualities as to soil, exposure to sunlight etc. The new pesticide is applied to 30 units while the old

pesticide to the remaining 30 units. Is there a reason to believe that the new pesticide is better than old pesticide if the mean number of kg's harvested per unit using new pesticide is 496.31 with S.D of 17.18 kg and for old pesticide the mean is 485.41 kg with S.D 14.73 kg. Test at 0.01 LOS.

→ ~~X_1, X_2 — Effect of new pesticide on rice production~~

→ X_1 - Amount of rice harvested using new pesticide
X_2 — " — old pesticide

→ H_0 :- $\mu_1 \leq \mu_2$

vs

H_1 :- $\mu_1 > \mu_2$ (New pesticide is better than the old pesticide)

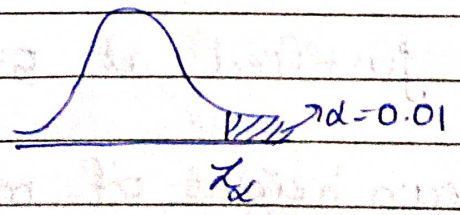
→ Testing Statistic,

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

→ Decision rule,

At $\alpha = 0.01$ LOS, we accept H_0 when

$Z < Z_\alpha$



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

$P(0<Z<Z_\alpha) = 0.490$

$\Rightarrow$ $\boxed{Z_\alpha = \cancel{1.96} \; 2.33}$

Page No.
Date

→ Numerical computation,

$n_1 = 30, \quad n_2 = 30, \quad \bar{x}_1 = 496.31, \quad \bar{x}_2 = 485.41$

$s_1 = 17.18, \quad s_2 = 14.73$

$$Z = (496.31 - 485.41) - (\cancel{\mu_1 - \mu_2}^{\,0}) \Big/ \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$Z = \frac{10.9}{4.1317}$$

$$\boxed{Z = 2.6381}$$

→ Conclusion,

$Z = 2.6381$ and $Z_\alpha = 2.33$

Since $Z \not< Z_\alpha$,

We reject H_0, i.e., new pesticide is better than the old pesticide.

3) To test the claim that men are taller than women, a survey was conducted resulting in following data.

Gender	Sample size	Mean height (in cm)	S.D in (cm)
Man	1600	172	6.3
Women	6400	170	6.4

Is the claim justified at 5% LOS?

→ X_1 – Mean height of men

X_2 – Mean height of women

Page No.
Date

→ H_0 :- $\mu_1 \leq \mu_2$

vs

H_1 :- $\mu_1 > \mu_2$ (Men are taller than women)

→ Testing statistic,

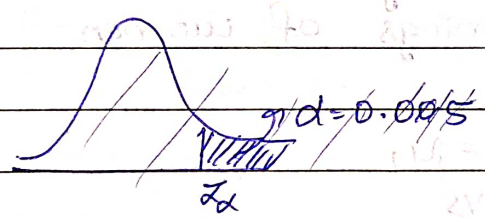
~~At $\alpha = 5\%$ LOS, we a~~

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

→ Decision rule,

At $\alpha = 5\%$ LOS, we accept H_0 when

$z < z_\alpha$



$0.05 = 0.5 - P(0 < z < z_\alpha)$

$P(0 < z < z_\alpha) = 0.450$

$\Rightarrow \boxed{z_\alpha = 1.64}$

→ Numerical computation,

$n_1 = 1600$, $n_2 = 6400$, $\bar{x}_1 = 172$, $\bar{x}_2 = 170$,

$s_1 = 6.3$, $s_2 = 6.4$

$$z = \frac{(172 - 170) - 0}{\sqrt{\frac{(6.3)^2}{1600} + \frac{(6.4)^2}{6400}}} = \frac{2}{0.176}$$

$\boxed{z = 11.32}$

Page No.
Date

→ Conclusion,

$$Z = 11.32 \text{ and } Z_\alpha = 1.64$$

Since, $Z \nless Z_\alpha$

We reject Ho, i.e., men are taller than women.

d) ~~If~~ In a random sample data, shows that 42 men earn an average 744.85 with S.D 397.7 while 32 women earn on an average 516.78 on a S.D 162.52. Test at 0.05 LOS whether avg. women income for men & women differ?

→ X_1 - Earnings of men
X_2 - Earnings of women

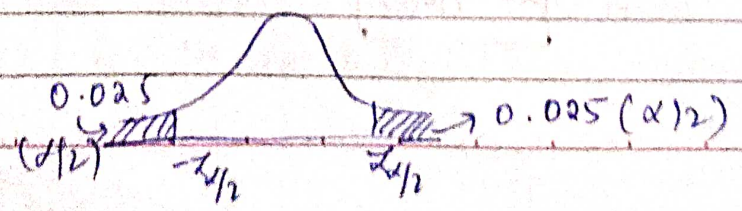
→ $H_0 :- \mu_1 = \mu_2$
vs
$H_1 :- \mu_1 \neq \mu_2$ (men earn more than women)

→ Testing statistic,

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

→ Decision rule,

At $\alpha = 0.05$ LOS, we accept Ho when $|Z| < Z_{\alpha/2}$



Page No.
Date

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

$$0.025 = 0.5 - P(0<Z<Z_{\alpha/2})$$

$$P(0<Z<Z_{\alpha/2}) = 0.475$$

$$\Rightarrow \boxed{Z_{\alpha/2} = 1.96}$$

→ Numerical computation,

$n_1 = 42$, $n_2 = 32$, $\bar{x}_1 = 744.85$, $\bar{x}_2 = \cancel{162.52}$ 516.78,

$S_1 = 397.7$, $S_2 = 162.52$

$$Z = \frac{(744.85 - 516.78) - (\cancel{\mu_1 - \mu_2})^{0}}{\sqrt{\frac{(397.7)^2}{42} + \frac{(162.52)^2}{32}}}$$

$$Z = \frac{228.07}{67.7587} = \underline{3.36}$$

→ Conclusion, $Z = 3.36$ and $Z_{\alpha/2} = 1.96$

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

Hence, we reject H_0 i.e., average income of men and women differ.

5) A company claims that alloying reduces resistance of electric wire by more than $0.05\,\Omega$. To test this claim ~~a~~ samples of standard wires and alloyed wires are tested yielding the following ~~wires~~ results

Type of wire	Sample size	Mean resistance (Ω)	S.D (Ω)
Standard	32	0.136	0.004
Alloyed	32	0.083	0.005

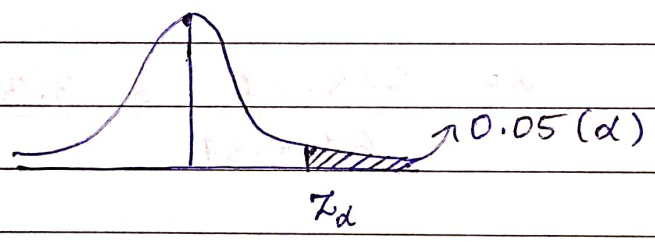
Date

→ X_1 - Resistance of standard wire
X_2 - Resistance of alloyed wire

→ H_0 :- $\mu_1 - \mu_2 \leq 0.05$
VS
H_1 :- $\mu_1 - \mu_2 > 0.05$

→ $$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

→ Decision rule,
At $\alpha = 0.05$ LOS, we accept H_0 when
$z < z_\alpha$



$0.05 = 0.5 - P(0 < z < z_\alpha)$
$P(0 < z < z_\alpha) = 0.45$

$\cancel{z_\alpha = 2.57}$
$\boxed{z_\alpha = 1.64}$

→ $n_1 = n_2 = 32$, $\bar{x}_1 = 0.136$, $\bar{x}_2 = 0.083$,
$S_1 = 0.004$, $S_2 = 0.005$

$$z = \frac{(0.136 - 0.083) - (0.05)}{\sqrt{\frac{(0.004)^2}{32} + (0.005)^2}} = \frac{0.003}{0.00132}$$

$z = 2.65$

→ Conclusion, $z = 2.65$ and $z_\alpha = 1.64$
$\Rightarrow z \nless z_\alpha$
Hence, we reject H_0 i.e., resistance decreases by more than 0.05 by using alloyed wire.

Date

Testing of hypothesis for single proportion

→ $X \sim B(n, p)$

$X \sim N(\mu = np, \sigma^2 = npq)$

→ deals with characteristic

→ $H_0 : P = P_0$

vs

$H_1 : P > P_0$

$P < P_0$

$P \neq P_0$

→ $z = \dfrac{p - E(p)}{S.E(p)} \sim N(0, 1)$

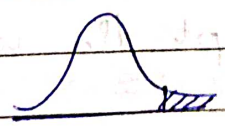
where, $E(p) = P$

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

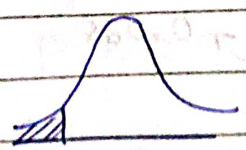
$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

where $p = \dfrac{x}{n}$

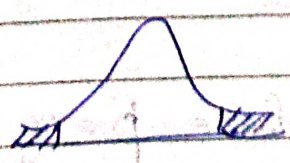
→ At $\alpha\%$, accept H_0 when



$z < z_\alpha$



$z > z_\alpha$



$|z| < z_{\alpha/2}$

Page No.
Date

Problems :-

1) If 48 out of 400 people in a rural area possess cellphones while 120 out of 500 in urban area. Can it be accepted the proportion of cellphones in rural & urban areas is same or not at $\alpha = 5\%$.

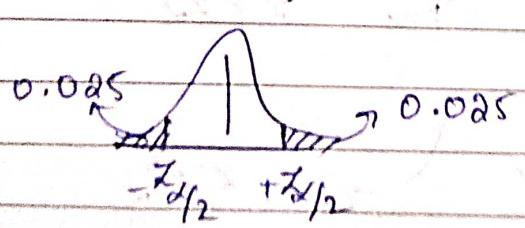
1) In a ~~so~~ random sample of 600 cars making a right turn at a certain traffic junctions, 157 drove in the wrong lane.
Test whether actually 30% of all drivers make this mistake or not at a given junction at $\alpha = 5\%$.

→ x - No. of cars driven into wrong lane

→ $H_0 :- P = 0.3$
vs
$H_1 :- P \neq 0.3$

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

→ At $\alpha = 5\%$, we accept H_0 when $|z| < z_{\alpha/2}$



$0.025 = 0.5 - P(0 < z < z_{\alpha/2})$
$P(0 < z < z_{\alpha/2}) = 0.475$

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

→ $$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$ where $p = \frac{x}{n} = \frac{157}{600} = 0.2617$

$P = 0.3 \Rightarrow Q = 0.7$
$n = 600$

$$z = \frac{0.2617 - 0.3}{\sqrt{\frac{0.3 \times 0.7}{600}}}$$

$$z = \frac{-0.03833}{0.01871}$$

$$z = -2.049$$

$$\Rightarrow |z| = 2.049$$

→ Conclusion,

$$|z| = 2.049 \ \& \ z_{\alpha/2} = 1.96$$

$$|z| \not< z_{\alpha/2}$$

Hence, we reject H_0 i.e, 30% of drivers do not make this mistake.

2) In a random sample of 200 persons suffering from headaches, 160 persons got cured by a drug. Can we accept the claim of the manufacturer that this drug cures 90% of the suffers. ($\alpha = 1\%$).

→ X — No. of persons getting cured by drug.

→ H_0 :- $P \geq 0.9$ (claim is true)

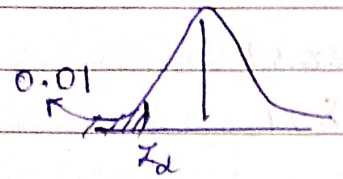
vs

H_1 :- $P < 0.9$ (claim is false)

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

→ At $\alpha = 1\%$, we accept H_0 when $z > z_\alpha$

$0.01 > 0.5 - P(0 < z < z_\alpha)$



$P(0 < z < z_\alpha) = 0.5 - 0.01 = 0.49$

$\Rightarrow \boxed{z_\alpha = -2.33}$

→ $n = 200$, $\bar{x} = 160$, $P = 0.9$, $Q = 0.1$, $p = \frac{x}{n} = \frac{160}{200}$

$p = 0.8$.

$$z = \frac{0.8 - 0.9}{\sqrt{\frac{0.9 \times 0.1}{200}}} = \frac{-0.1}{0.02121}$$

$$\boxed{z = -4.714}$$

→ Conclusion, $z = -4.714$ and $z_\alpha = -2.33$

$\Rightarrow z \not> z_\alpha$

Hence, we reject H_0 i.e., claim is false.

3) In a survey of injection drug users in a large city, it is found that 18 out of 423 were HIV positive, we wish to know if it can be computed that fewer than 5% of injection drug users in sample proportion are HIV positive.

→ X – No. of HIV drug users

Page No.
Date

→ H_0 :- $P \geq 0.05$

vs

H_1 :- $P < 0.05$

0.50
0.05
0.45

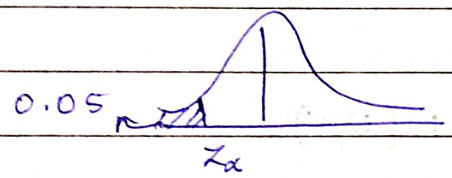
→ $$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

→ At $\alpha = 5\%$, we accept H_0 when $z > z_\alpha$

$0.05 = 0.5 - P(0 < z < z_\alpha)$

$P(0 < z < z_\alpha) = \cancel{0.495}$ 0.45

$\Rightarrow \boxed{z_\alpha = -1.64}$



1.00
0.05
0.95

→ $x = 18$, $n = 423$, $p = \frac{x}{n} = \frac{18}{423}$, $P = 0.05$, $Q = 0.95$

$$z = \frac{18/423 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{423}}} = \frac{-0.0074}{10.59 \times 10^{-3}}$$

$z = -0.698 \cancel{\times 10} \sim -0.7$

→ Conclusion,

$z = -0.698$ and $z_\alpha = -1.64$

$\Rightarrow z > z_\alpha$

Hence, we accept H_0 i.e., not fewer than 5% of injection drug users in sample proportion are HIV positive.

Page No.
Date

4) In a random sample of 120 tractors produced by a company, 47 are defected. Is the claim by the company that atmost 30% of tractors are defective tenable (justifiable)? ($\alpha = 0.05$).

→ X – No. of defected tractors

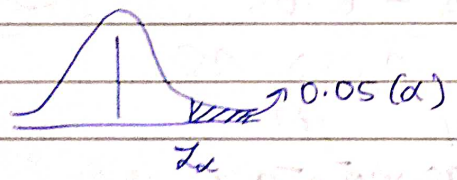
→ $H_0 : P \leq 0.3$

vs

$H_1 : P > 0.3$

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

→ At $\alpha = 0.05$ LOS, we accept H_0 when $Z < Z_\alpha$



$0.05 = 0.5 - P(0 < Z < Z_\alpha)$

$P(0 < Z < Z_\alpha) = 0.450$

$\Rightarrow \boxed{Z_\alpha = 1.64}$.

→ $x = 47$, $n = 120$, $P = 0.3$

$\Rightarrow p = \frac{x}{n} = 0.392$, $Q = 0.7$

$$Z = \frac{0.392 - 0.3}{\sqrt{\frac{0.3 \times 0.7}{120}}} = \frac{0.092}{0.0418} = 2.19$$

$\Rightarrow \boxed{Z = 2.19}$.

Page No.
Date

→ Conclusion,

$$z = 2.19 \text{ and } z_\alpha = 1.64$$

$$\Rightarrow z \not< z_\alpha$$

Hence, we reject H_0 i.e., the defective tractors produced by the company is not atmost 30% rather greater than 30%

5) In a random sample of 400 persons of large population, 120 are females. Can it be said that males & females are in the ratio 5:3 in the population? ($\alpha = 1\%$).

→ X – No. of females

→ H_0 :- $P = 3/8$

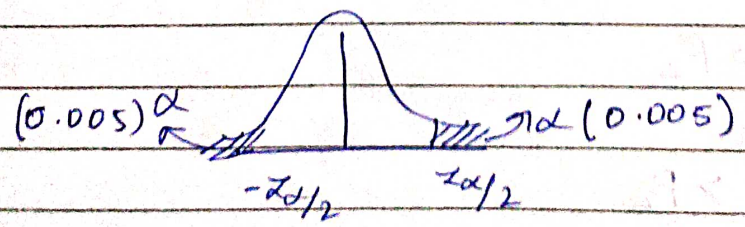
vs

H_1 :- $P \neq 3/8$

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

→ At $\alpha = 1\%$ LOS, we accept H_0 when $|z| < z_{\alpha/2}$

$$0.005 = 0.5 - P(0 < z < z_{\alpha/2})$$



$$P(0 < z < z_{\alpha/2}) = 0.495$$

$$\Rightarrow \boxed{z_{\alpha/2} = 2.57}$$

Date

$\rightarrow x = 120, \; n = 400, \; P = 3/8, \; Q = 5/8$

$\Rightarrow p = \frac{120}{400} = 0.300$

$$z = \frac{0.3 - 0.375}{\sqrt{\frac{0.375 \times 0.625}{400}}} = \frac{-0.075}{0.0242}$$

$$\boxed{z = -3.098}$$

$\rightarrow$ Conclusion,

$z = -3.098$ and $z_{\alpha/2} = 2.57$

$\Rightarrow |z| \nless z_{\alpha/2}$

Hence, we reject H_0 i.e., ratio of males to females is not 5:3.

Test of significance for difference of proportions.

$\rightarrow X_1 \sim B(n_1, P_1)$
$X_2 \sim B(n_2, P_2)$

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

Page No.
Date

$$\rightarrow \quad z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

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

or $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

and $Q = 1 - P$

Problems:-

1) If 57 out of 150 patients suffering with certain disease are cured by allopathy and 33 out of 100 patients with same disease are cured by homeopathy. Is there a reason to believe that allopathy is better than homeopathy at 5% LOS?

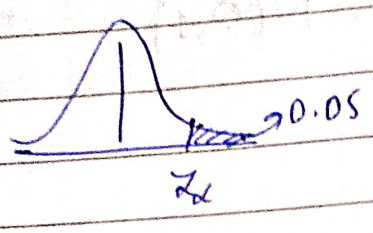
→ X_1 - No. of patients cured by allopathy
X_2 - No. of patients cured by homeopathy

→ H_0 :- $P_1 \leq P_2$ (allopathy is not better than homeopathy)
VS
H_1 :- $P_1 > P_2$ (— " — better — " —)

$$\rightarrow \quad z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

→ ~~We accept~~ At $\alpha = 0.05$ LOS, we accept H_0 when $z < z_\alpha$

$$0.05 = 0.5 - P(0<z<z_\alpha)$$
$$P(0<z<z_\alpha) = 0.45$$
$$\Rightarrow \boxed{z_\alpha = 1.64}$$



→ $x_1 = 57$, $x_2 = 33$ ⇒ $p_1 = \frac{x_1}{n_1} = 0.380$
$n_1 = 150$, $n_2 = 100$

$$p_2 = \frac{x_2}{n_2} = 0.33$$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = 0.360 \quad , \quad Q = 0.64$$

$$z = \frac{(0.380 - 0.33) - 0}{\sqrt{0.36 \times 0.64\left(\frac{1}{150} + \frac{1}{100}\right)}}$$

$$z = \frac{0.050}{0.061} = 0.807$$

→ Conclusion,

$z = 0.8$ and $z_\alpha = 1.64$
$\Rightarrow z < z_\alpha$

Hence we accept H_0 i.e., no. of people getting cured by allopathy is not more than patients getting cured by homeopathy.

2) Out of the 2 vending machines at a superbazaar, the first machine fails to work 13 times in 250 trials, and second machine fails to work 7 times

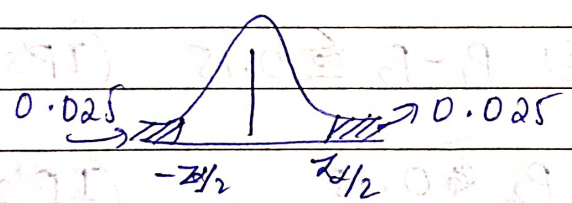
in 250 trials. Test at 0.05 LOS whether the difference b/w corresponding ~~mach~~ sample proportions is significant.

→ X_1 – No. of times vending machine 1 fails to work.
X_2 – —— " —— 2 fails to work.

→ H_0 :- $P_1 = P_2$ (Data is insignificant)
vs
H_1 :- $P_1 \neq P_2$ (Data is significant)

→ $$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

→ At $\alpha = 0.05$ LOS, we accept H_0 when $|z| < z_{\alpha/2}$



$0.025 = 0.5 - P(0 < z < z_{\alpha/2})$
$P(0 < z < z_{\alpha/2}) = 0.475$

$\Rightarrow \boxed{z_{\alpha/2} = 1.96}$

→ $x_1 = 13$, $x_2 = 7$, $n_1 = n_2 = 250$, $P = \frac{x_1 + x_2}{n_1 + n_2} = 0.040$
$\Rightarrow p_1 = \frac{x_1}{n_1} = 0.052$ $p_2 = \frac{x_2}{n_2} = 0.028$, $\Rightarrow Q = 0.96$

$$z = \frac{(0.052 - 0.028) - 0}{\sqrt{0.04 \times 0.96 \times \left(\frac{1}{500}\right)}} = \frac{0.024}{\cancel{0.0087}\ 0.0124}$$

$\boxed{z = 1.93}$

Page No.
Date

→ Conclusion, $z = 1.93$ and $z_{\alpha/2} = 1.96$

$$\Rightarrow |z| < z_{\alpha/2}$$

Hence we accept H_0 i.e., data is insignificant.

3) A question in a true-false quiz is considered to be smart if it discriminates b/w intelligent person (IP) & an average person (AP). Suppose 205 out of 250 IP's and 137 out of 250 AP's answer a quiz question correctly. Test at 0.01 LOS whether for the given question the proportion of correct answers can be expected to be atleast 15% higher among IP's than among AP's.

→ X_1 - No. of correctly answered questions by IP's.
X_2 - —"— AP's

→ H_0 :- $P_1 \geq P_1 - P_2 \leq 0.15$ (IP's > AP's)
vs
H_1 :- $P_1 - P_2 > 0.15$ (IP's < AP's)

→ $$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

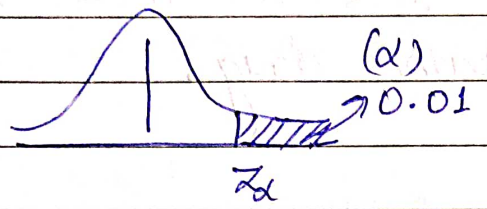
→ At $\alpha = 0.01$ LOS, We accept H_0 when

Page No.
Date

$$\rightarrow \quad z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1)$$

$\rightarrow$ At $\alpha = 0.01$ LOS, we accept H_0 when $z < z_\alpha$

$$0.01 = 0.5 - P(0 < z < z_\alpha)$$



$$P(0 < z < z_\alpha) = 0.490$$

$$\Rightarrow \boxed{z_\alpha = 2.33}$$

$\rightarrow$ $x_1 = 205$, $n_1 = 250$, $p_1 = \frac{x_1}{n_1} = 0.820$, $q_1 = 0.180$

$x_2 = 137$, $n_2 = 250$, $p_2 = \frac{x_2}{n_2} = 0.548$, $q_2 = 0.452$

$$z = \frac{(0.820 - 0.548) - 0.15}{\sqrt{\frac{(0.82 \times 0.18) + (0.548 \times 0.452)}{250}}}$$

$$z = \frac{0.122}{0.0398}$$

$$\boxed{z = 3.065}$$

$\rightarrow$ Conclusion,

$z = 3.065$ and $z_\alpha = 2.33$

$\Rightarrow z \not< z_\alpha$

Hence, we reject H_0 i.e., proportion of correct answers given by IP's ~~are~~ is greater than AP's.

Page No.
Date

24) In a study of effect of drugs on cancer, 2 groups of 80 such patients were considered. One group was treated with allopathy while other by homeopathy. It was observed that 23 in first group & 41 in 2nd group were cured. At 0.05 LOS test whether the true % of patients cure is atleast 8% less for those who were treated by homeopathic drug.

Page No.
Date

Testing equality of Variances :-

→ $X_1 \sim N(\mu_1, \sigma_1^2)$
$X_2 \sim N(\mu_2, \sigma_2^2)$

→ $H_0 :- \sigma_1^2 = \sigma_2^2$
vs
$H_1 :- \sigma_1^2 \neq \sigma_2^2$

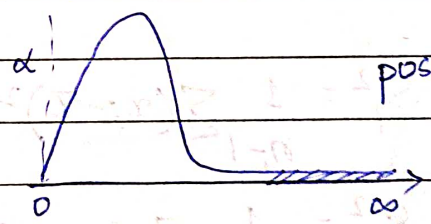
→ $F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1, \alpha}$

Note :- Numerator > Denominator

$\frac{S_2^2}{S_1^2} \sim F_{n_2-1, n_1-1, \alpha}$

F - Snedecor's distribution

→ Continous prob. distribution
→ pdf is expressed in terms of β function.
→ $0 < \alpha < \infty$
→



positively skewnes

→ At α%, accept H_0 when $F < F_{n_1-1, n_2-1, \alpha}$

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

Problems :-

1) Can we conclude that 2 population variances are equal for the following data of post graduates passed out from the state and private universities?

In calculator, $\Sigma x = S$

$S_x = S$

Page No.
Date

Data:-

State	8350,	8260,	8130,	8340,	8070	
Pvt. Uni.	7890,	8140,	7900,	7950,	7840,	7920

→ X_1 - Passed out post graduates from State Universities
X_2 - —— " —— private universities

→ H_0 :- $\sigma_1^2 = \sigma_2^2$

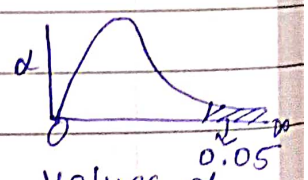
vs

H_1 :- $\sigma_1^2 \neq \sigma_2^2$

→ $$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1, \alpha}$$

→ At $\alpha = 5\%$, accept H_0 when $F < F_{n_1-1, n_2-1, \alpha}$

$\underline{F_{4,5,0.05} = 5.19}$



Values of $F_{0.05}(\nu_1, \nu_2)$

→ $$S_1^2 = \frac{1}{n_1-1}\sum(x_i-\bar{x})^2 = 15750$$

$$S_2^2 = \frac{1}{n_2-1}\sum(x_i-\bar{x})^2 = 10920$$

$S_1^2 > S_2^2$

$$F = \frac{S_1^2}{S_2^2} = 1.4423 \sim \underline{1.44}$$

→ Since $F < F_{4,5,0.05}$

We accept H_0 i.e., passed out p.g from State & private universities are equal.

Page No.
Date

2) 2 varieties of tomato were grown in field plots. Yield (kg's) data are as follows.

Variety One 1.4 1.1 1.8 1.2, 1.2, 1.8, 0.8, 1.0, 1.2 1.6

Variety 2 1.0 1.2 1.0 1.6 1.7 0.7 0.8 1.3

Test whether the varieties differ in yield.

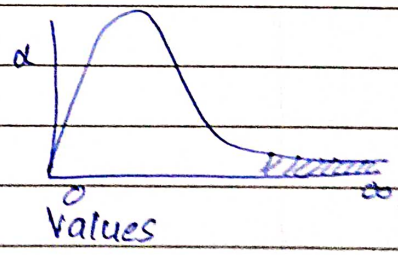
→ X_1 - Yield of variety 1 tomatoes.
X_2 - Yield of variety 2 tomatoes.

→ $H_0 :- \sigma_1^2 = \sigma_2^2$ vs $H_1 :- \sigma_1^2 \neq \sigma_2^2$
(Yields are same) (Yields are not same)

→ $F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1, \alpha}$

→ At $\alpha = 5\%$, accept H_0 when $F < F_{n_1-1, n_2-1, \alpha}$

$F_{7,9,0.05} = \cancel{3.68}$
3.29



→ $S_1^2 = \frac{1}{n_1-1} \sum_1 (x_i - \bar{x})^2 = 0.1112$

$S_2^2 = \frac{1}{n_2-1} \sum_1 (x_i - \bar{x})^2 = 0.1284$

$S_1^2 < S_2^2$

$F = \frac{S_1^2}{S_2^2} = \frac{0.112}{0.1284} = \cancel{0.8747}$ 1.14

→ Since $F < F_{9,7,0.05}$, we accept H_0 i.e., yields of both the varieties of tomatoes are equal.

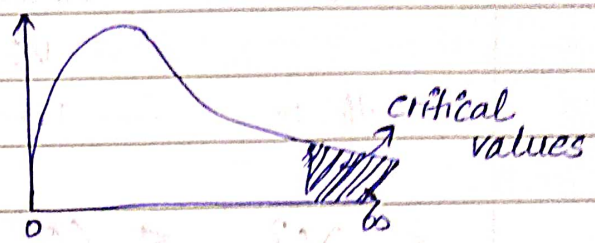
$\sigma_1^2 = \sigma_2^2$ (insignificant)
$\sigma_1^2 \neq \sigma_2^2$ (significant)

Page No.
Date

3) In one sample of 8 observations, the sum of squares of deviation of sample value from sample mean was 84.4 and in another sample of 10 observations, it was 102.6. Test whether variation is significant at 5% LOS.

Chi Square Distribution (χ^2)

→ Non-parametric test
→ Distribution free test
→ Positively Skewed
→ Deals with attributes
→ Applications :-
  * Goodness of fit
  * Independence of attributes
  * Homogenity of data



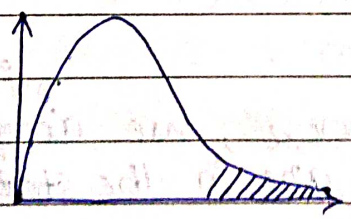
Testing of Hypothesis for goodness of fit :-

→ H_0 :- Distribution is a good fit.
vs
H_1 :- Distribution is not a good fit.

→ $$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-k}$$ → (degree's of freedom)

n - no. of observations
k - no. of constraints

→ At $\alpha\%$, we accept H_0 when



$$\chi^2 < \chi^2_{n-k,\alpha}$$
(table values)

Problems :-

1) A sample analysis of examination results of 500 students was made, it was found that 220 students had failed, 170 had secured 3rd class, 90 had secured 2nd class and 20 secured 1st class. Do these results figures

supports the general examination results which is in the ratio 4:3:2:1 respectively.

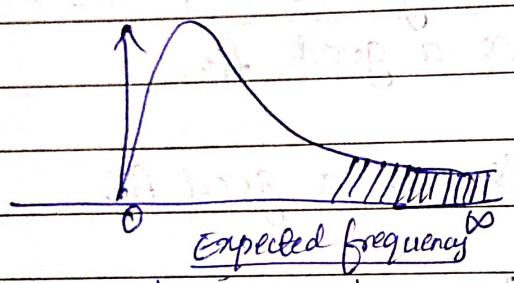
→ H_0 :- Exam results are in the stated ratio.

VS

H_1 :- —"— not —"—.

→ $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-k}$

→ At $\alpha = 5\%$ LOS, we accept H_0 when $\chi^2 < \chi^2_{n-k,\alpha}$



$\chi^2_{n-k,\alpha}$; $n = 4$

$k = 1$ [$\Sigma O = \Sigma E$] Default

$\chi^2_{4-1,0.05}$ $\boxed{= 7.815}$

→	O	E	χ^2
Failed	220	$\frac{4}{10} \times 500 = 200$	~~0.1000~~ 2.000
3rd	170	$\frac{3}{10} \times 500$ 150	~~0.1833~~ 2.6667
2nd	90	100	1
1st	20	50	18
$n=4$	500	500	$\boxed{\chi^2 = 23.6667}$

→ Since, $\chi^2 \nless \chi^2_{n-k,\alpha}$, we ~~accept~~ reject H_0 i.e., exam results are in the stated ratio.

2) In a Mendelian exp., 4 kinds of plants are expected to occur in the proportion 9:3:3:1. The observed frequencies are 891 are round & yellow, 316 wrinkled and yellow, 290 round & green and 119 wrinkled & green. Find the χ^2 value and examine the correspondence b/w

theory & experiment.

→ H_0 :- There is a correspondence in b/w theory & exp.

vs

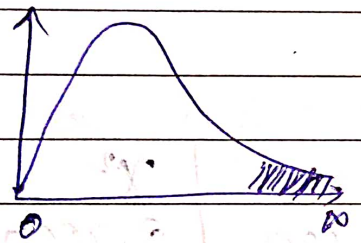
H_1 :- —"— not —"—

→ $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-k,\alpha}$

→ At $\alpha = 5\%$ LOS, we accept H_0 when $\chi^2 < \chi^2_{n-k,\alpha}$

Here $n=4$, $k=1$

$\therefore \boxed{\chi^2_{3,0.05} = 7.815}$



	O	E	χ^2
→ R&Y	891	$\frac{9}{16} \times 1616 = 909$	0.3564
W&Y	316	$\frac{3}{16} \times 1616 = 303$	0.5577
R&G	290	$\frac{3}{16} \times 1616 = 303$	0.5577
W&G	119	$\frac{1}{16} \times 1616 = 101$	3.2079
$n=4$	1616	1616	$\chi^2 = 4.6797$

→ Since $\chi^2 < \chi^2_{3,0.05}$, we accept H_0 i.e., there is a correspondence in b/w theory and experiment.

3) It is believed that the prop. of people with blood types A, B, O and AB in a population are respectively 0.4, 0.2, 0.3 and 0.1. When 400 randomly picked people were examined the no. of persons with these blood types was observed to be 148, 96, 106 and 50 respectively. Test the hypothesis that these data were out the stated belief.

→ H_0 :- Blood types are in the stated ratio.

vs

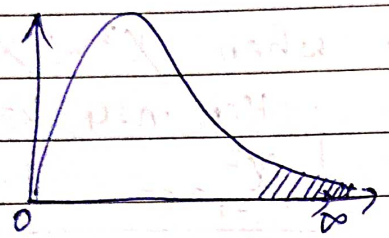
H_1 :- —"— not —"—.

→ $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{n-k,\alpha}$

→ At $\alpha = 5\%$ LOS, we accept H_0 when $\chi^2 < \chi^2_{n-k,\alpha}$

Here $n=4$, $k=1$

$\boxed{\chi^2_{3,0.05} = 7.815}$



Blood Types	O	E	χ^2
→ A	148	$0.4 \times 400 = 160$	0.9000
B	96	$0.2 \times 400 = 80$	3.2
O	106	$0.3 \times 400 = 120$	1.6333
AB	50	$0.1 \times 400 = 40$	2.5
$n=4$	400	400	$8.2333 = \chi^2$

→ Since, $\chi^2 \nless \chi^2_{3,0.05}$, we reject H_0 i.e., the blood types are not in the stated ratio.

4) Fit a poisson distn to the following data & test for the goodness of fit at $\alpha = 5\%$.

X	0	1	2	3	4
F	419	352	154	56	19

→ $X \sim P(\lambda)$

$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2, \ldots$

Page No.
Date

→ H_0 :- Poisson distribution is a good fit

vs

H_1 :- —"— not —"—.

→ x	f	fx	$E = Np(x)$
0	~~382~~ 419	0	~~366.1~~ 404.9 ~ 405
1	~~154~~ 352	352	366.1 ~ 367
2	154	308	~~183.0~~ 165.5 ~ 166
3	56	168	~~61.0~~ 49.85 ~ 50
4	19	76	~~15.28~~ 11.26 ~ 12
$n=5$	$N=1000$	904	1000

Here λ is unknown, replace with its unbiased mean

$$\hat{\lambda} = \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{904}{1000} = 0.904$$

$p(0) = e^{-0.904}(0.904)^0 = \cancel{0.3661}\ 0.4049$

$p(1) = 0.3661$

$p(2) = \cancel{0.1830} = 0.1655$

$p(3) = 0.0610 = 0.04985$

$p(4) = 0.01525 = 0.01126$

$$\chi^2 = \frac{(O-E)^2}{E}$$

Page No.
Date

5) It is desired to test whether the no. of γ-rays emitted/sec by a certain radioactive substance is a random variable having a Poisson distr with mean 2.4. Use the following data obtain for 300 1-second intervals to test this null hypothesis.

No. of γ-rays	0	1	2	3	4	5	6	7 or more
f	19	48	66	74	44	35	10	4

→ X – No. of γ-rays emitted/sec by a certain radioactive substance.

→ H_0 :- Poisson distr is a good fit.

vs

H_1 :- —"— not —"—.

→ $X \sim P(x)$

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, 2, 3, \ldots \qquad \lambda = \mu = 2.4$$

→

x	f or O	$p(x)$	$Np(x) = E$	$\chi^2 = \frac{(O-E)^2}{E}$
0	19	0.0907	27.21 ~ 27	2.3704
1	48	0.2177	65.31 ~ 65	4.4461
2	66	0.2613	78.39 ~ 78	1.8461
3	74	0.2090	62.70 ~ 63	1.9206
4	44	0.1254	37.62 ~ 38	0.9474
5	35	0.060 a	18 ~ 18	16.0555
6	10	0.024	7.2 ~ 7	1.2857
7 or more	4	0.0119 (1-a)	3.57 ~ 4	0
	$N = 300$		300	$\chi^2 = 28.87$

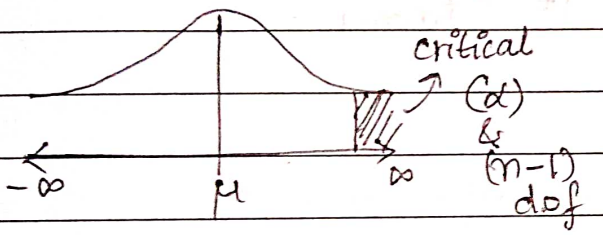
Page No.
Date

Testing of Hypothesis for single mean (Small Sample) :-

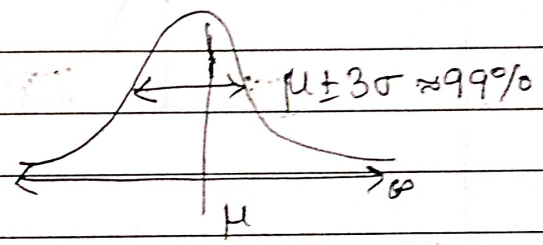
$$\frac{\bar{x} - E(\bar{x})}{S.E(\bar{x})} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

t - distribution / Student's t distribution.

→ Continous prob. distribution.
→ Similar to normal distribution



t- curve



normal dist'

Degrees of freedom : The no. of values in the final calculation of a statistic, that are free to vary.

→ $X \sim N(\mu, \sigma^2)$
Samples are drawn randomly.

→ $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$
$\mu < \mu_0$
$\mu \neq \mu_0$

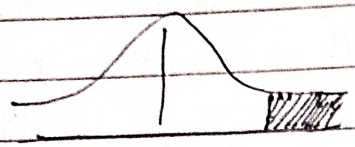
→ Test Statistic,

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim t_{\alpha, n-1}$$

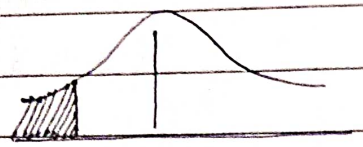
$$\hat{\sigma} = S = \sqrt{\frac{1}{n-1} \sum_i (x - \bar{x})^2}$$

In calci, $\sigma = s$
$s = S$

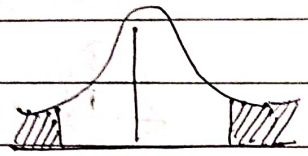
→ Decision rule,
At $\alpha\%$, accept H_0 when



$t < t_{\alpha, n-1}$



$t > -t_{\alpha, n-1}$



$|t| < t_{\alpha, n-1}$

Problem :-

1) An ambulance service company claims that on an average it takes 20 minutes between a call for an ambulance and the patients arrival at the hospital. If in 6 calls, the time taken (b/w a call & arrival at hospital) are 27, 18, 26, 15, 20 and 32. Can the company's claim be accepted at 0.05 LOS?

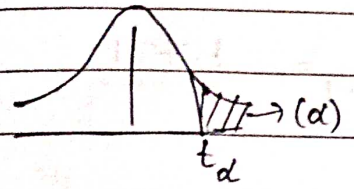
→ X – Time taken b/w a call and the arrival of patient at the hospital.

→ H_0 :- $\mu \leq 20$ (Company's claim is true)
vs
H_1 :- $\mu > 20$ (Company's claim is false)

→ $t = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{\alpha, n-1}$

Page No.
Date

→ At $\alpha = 5\%$ LOS, accept H_0 when



$t_\alpha = 2.015$
$0.05, 5$

→ $\bar{x} = \frac{\Sigma x}{n} = 23$

$S = \sqrt{\frac{1}{(n-1)} \Sigma_i (x-\bar{x})^2} = 6.3875$

$t = 1.15$

→ $t < t_\alpha$, Hence we accept H_0.
∴ The Company's claim is true.

2) The mean weekly sales of the chocolate bar in a candy stores was 140.3 bars per store. After an advertise-ment campaign, the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a standard deviation of 17.2. was the advertisement campaign unsuccessful?

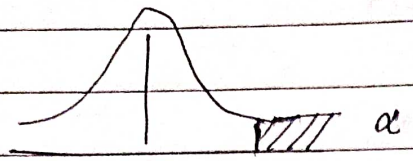
→ X – weekly sales of the chocolate bar

→ H_0 :- $\mu = 140.3$ (Unsuccessful)
vs
H_1 :- $\mu > 140.3$ (Successful)

→ $t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{\alpha, n-1}$

→ At $\alpha = 5\%$, we accept H_0 when

$$t_{0.05,\,22-1} = 1.721$$



→ $\bar{x} = \cancel{28}\ 153.7$, $S = 17.2$

$$t = \frac{\cancel{28}\ 153.7 - 140.3}{17.2/\sqrt{22}}$$

$$\boxed{t = 3.6542}$$

→ $t \not< t_\alpha \Rightarrow$ We reject H_0.

Hence, the campaigning was successful.

3) The following are the systolic B.P of 12 patients undergoing drug therapy for hypertension.
183, 152, 178, 157, 194, 163, 144, 114, 178, 152, 118, 158.
Can we conclude on the basis of these data that the popⁿ mean is less than 165 using $\alpha = 0.05$ LOS)

→ X - Systolic BP of a patient

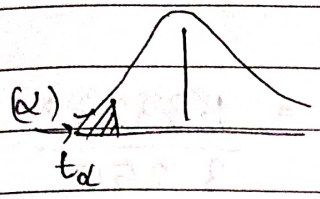
→ $H_0: \mu \geq 165$ (Popⁿ mean is more than 165)

vs

$H_1: \mu < 165$ (popⁿ mean is less than 165)

Page No.
Date

→ At $\alpha = 0.05$ LOS, we accept H_0 when $t > -t_{\alpha, n-1}$



$$t_{0.05, 11} = -1.796$$

→ $\bar{x} = \frac{\Sigma x}{n} = 157.5833$

$$S = \sqrt{\frac{1}{n-1} \Sigma_i (x - \bar{x})} = 24.3999$$

$$t = \frac{157.5833 - 165}{24.3999/\sqrt{12}} = -1.0529$$

→ $t > -t_\alpha$, ⇒ we accept H_0

Hence, population mean is not less than 165.

4) Test the hypothesis that the average content of containers of a particular lubricant is 10 ltrs. If the contents of a random sample of 10 containers are 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, 10.2, 9.7, 9.8 ltrs ($\alpha = 1\%$)

→ x – Content of containers of a particular lubricant.

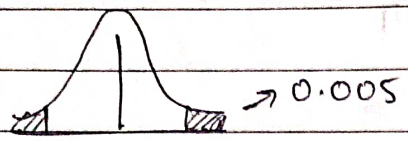
→ H_0 :– $\mu = 10$

vs

H_1 :– $\mu \neq 10$

→ $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{\alpha, n-1}$

→ At $\alpha = 1\%$ LOS, we accept H_0 when $|t| < t_{0.05,9}$



$t_{\alpha, n-1} = t_{0.005,9} = \cancel{1.83333}$ 3.250

→ $\bar{x} = 10.06$

$s = \sqrt{\frac{1}{n-1}\sum_i (x-\bar{x})} = 0.2458$

$t = \frac{10.06 - 10}{0.2458} \times \sqrt{10} = 0.77191$

→ $|t| < t_\alpha \Rightarrow$ we ~~reject~~ accept H_0

Hence, average content of containers of lubricant is ~~not~~ 10 ltrs.

Page No.
Date

Testing of difference of mean for small samples :-

→ Comparative study of 2 population mean.

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

→ Samples are independent

→ H_0 :- $\mu_1 - \mu_2 = S$

vs

H_1 :- $\mu_1 - \mu_2 \neq S$

$\mu_1 - \mu_2 < S$

$\mu_1 - \mu_2 > S$

→ Decision rule,

Accept H_0, when

$|t| < t_{\alpha/2, n_1+n_2-2}$

$t > t_{\alpha, n_1+n_2-2}$

$t < t_{\alpha, n_1+n_2-2}$

→ $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{\alpha, n_1+n_2-2}$$

S^2 - Pooled variance

$$= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$S^2 = \frac{1}{n-1}\sum(x-\bar{x})^2, \quad s^2 = \frac{1}{n}\sum(x-\bar{x})^2$$

→ ~~Acc~~

$$\Rightarrow \boxed{(n-1)S^2 = ns^2}$$

Page No.
Date

Paired t test :-

→ $X_1 \sim N(\mu_1, \sigma_1^2)$
$X_2 \sim N(\mu_2, \sigma_2^2)$

→ Samples are dependent.

$$d_i = X_1 - X_2$$

$$d \sim N(\mu_d, \sigma_d^2)$$

→ H_0 :- $\mu_1 - \mu_2 = 0$ or $\mu_d = 0$ → Accept H_0 when

vs

H_1 :- $\mu_1 \neq \mu_2$ or $\mu_d \neq 0$ $\quad |t| < t_{\alpha/2, n-1}$

$\mu_1 - \mu_2 < 0$ or $\mu_d < 0$ $\quad t > t_{\alpha, n-1}$

$\mu_1 - \mu_2 > 0$ or $\mu_d > 0$ $\quad t < t_{\alpha, n-1}$

→ $$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}} \sim t_{\alpha, n-1}$$

$$\bar{d} = \frac{\Sigma d}{n} \qquad S_d = \frac{1}{(n-1)} \Sigma (d_i - \bar{d})^2$$

Problems :-

1) In a study of usefulness of yoga in weight reduction, a random sample of 8 persons undergoing yoga were examined of their weight before and after yoga with the following results.

Page No.
Date

Weight before	209	178	169	212	180	192	158	180
Weight after	196	171	170	207	177	190	159	180

Test whether yoga is useful in weight reduction at $\alpha = 1\%$.

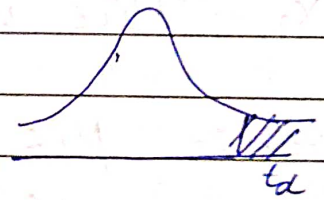
→ X_1 – weight before the yoga
X_2 – weight after the yoga.

→ H_0 :- $\mu_1 \le \mu_2$ vs or $\mu_d \le 0$ vs (Yoga is not useful)

H_1 :- $\mu_1 > \mu_2$ $\mu_d > 0$ (Yoga is useful)

→ $t = \dfrac{\bar{d} - \mu_d}{S_d/\sqrt{n}} \sim t_{\alpha, n-1}$

→ At $\alpha = 1\%$, accept H_0 when
$t < t_{\alpha, n-1}$
$t_{0.01, 7} = 2.998$



→ $x_1 \rightarrow 209, 178, \ldots, 180$
$x_2 \rightarrow 196, 171, \ldots, 180$
$d \rightarrow 13, 7, -1, 5, 3, 2, -1, 0$

$\bar{d} = \dfrac{\Sigma d}{n} = 3.5$

$S_d = 4.7809$

$\Rightarrow t = \dfrac{3.5 - 0}{4.7809/\sqrt{8}}$

$\boxed{t = 2.0115}$

→ Since, $t < t_{0.01,7}$, and hence we accept H_0 i.e., Yoga is not useful in weight reduction.

2) Out of random sample of 9 mice, suffering with a disease, 5 mice were treated with a new serum while the remaining were not treated. From the time commencement of exp., the following are the survival times.

Treatment	2.1	5.3	1.4	4.6	0.9
No Treatment	1.9	0.5	2.8	3.1	

Test whether the serum treatment is effective in curing the disease at 5% LOS, assuming that the two distributions are normally distributed with equal variances ($\sigma_1^2 = \sigma_2^2$).

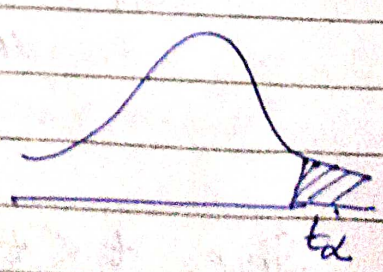
→ X_1 — Survival time of Mice with serum treatment
X_2 — " Mice without serum treatment

→ H_0 :- $\mu_1 \leq \mu_2$ — Serum treatment is ineffective
vs
H_1 :- $\mu_1 > \mu_2$ — Serum treatment is effective

→ $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{\alpha, n_1+n_2-2}$$

→ We accept H_0 when
$t < t_{\alpha, n_1+n_2-2}$

$t_{0.05,(5+4-2)} = t_{0.05,7} = 1.895$



Page No.
Date

→ $n_1 = 5$, $\bar{x}_1 = \frac{\Sigma x_1}{n_1} = 2.86$, $\bar{x}_2 = \frac{\Sigma x_2}{n_2} = 2.075$
$n_2 = 4$

$$t = \frac{(2.86 - 2.075) - 0}{1.07\sqrt{\frac{1}{5} + \frac{1}{4}}}$$

$$S^2 = \frac{(4)(1.9705)^2 + 3(1.1673)^2}{7}$$

$$S^2 = \cancel{1.6263}\ 2.8027$$

$$\boxed{\cancel{t = 0.3153}}$$

$$\boxed{t = 0.7007}$$

→ Since $t < t_{0.05,7}$, we accept H_0.

Hence, serum treatment is ineffective.

3) A study is conducted to determine whether the wear of material A exceeds that of B by more than 2 units. If test of 12 pieces of material A yielded a mean wear of 85 units & S.D of 4 while test of 10 pieces of material B yielded a mean of 81 and S.D of 5 units, what conclusion can be drawn at 5% LOS. Assume that populations are approximately normally distributed with equal variances ($\sigma_1^2 = \sigma_2^2$)

→ X_1 – wear of material A
X_2 – wear of material B

→ H_0 :– $\mu_1 - \mu_2 \leq 2$ (wear of A doesn't exceed wear of B)
vs
H_1 :– $\mu_1 - \mu_2 > 2$ (–"– exceeds –"–)

Page No.
Date

→ $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{\alpha, n_1+n_2-2}$

$\bar{x}_1 = 85$, $S_1 = 4$, $n_1 = 12$
$\bar{x}_2 = 81$, $S_2 = 5$, $n_2 = 10$

$t = \dfrac{(85-81) - 2}{4.4777\sqrt{\frac{1}{12} + \frac{1}{10}}}$

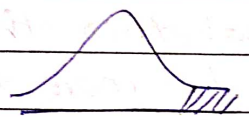
$S^2 = \dfrac{11(16) + 9(25)}{20}$

$S^2 = 20.05$

$S = 4.4777$

$\boxed{t = 1.0432}$.

→ At $\alpha = 5\%$, we accept H_0 when $t < t_{\alpha, n_1+n_2-2}$



$t_{0.05, 20} = \underline{1.725}$

→ Since, $t < t_{0.05, 20}$, we accept H_0.

Hence, wear of material A does not exceeds wear of material B.

4) The pulsality index (P.I) of 11 patients before and after contracting a disease is given below. Test at 0.05 level of significance whether there is a significant increase of the mean of P.I. values.

Before	0.4	0.45	0.44	0.54	0.48	0.62	0.48	0.6	0.45	0.46	0.35
After	0.5	0.6	0.57	0.65	0.63	0.78	0.63	0.8	0.69	0.62	0.68

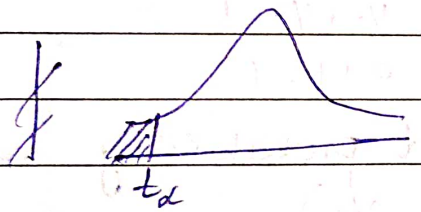
Page No.
Date

→ X_1 – PI of a patient by contracting a disease.
X_2 – —II— after —II—.

→ H_0 :- $\mu_d \geq 0$ (PI doesn't increase after contracting the disease)
vs
H_1 :- $\mu_d < 0$ (—II— does increase —II—)

→ $t = \dfrac{\bar{d} - \mu_d}{S_d/\sqrt{n}} \sim t_{\alpha, n-1}$

→ At $\alpha = 5\%$ LOS, we accept H_0 when $t > t_{\alpha, n-1}$



$-t_{0.05, 11-1}$

$-\boxed{t_{0.05, 10} = -1.812}$

→ $d = -0.1, -0.18, -0.13, -0.11, -0.15, -0.16, -0.15, -0.2, -0.24, -0.16, -0.33$

$\bar{d} = \dfrac{\Sigma d}{n} = -0.1709$

$S = 0.0654$

$$t = \frac{-0.1709 - 0}{0.0654/\sqrt{11}} = -8.6668$$

→ Since, $t \not> t_{0.05, 10}$, we reject H_0.

Hence, P.I does increase after contracting the disease.