unit - IV

Sample estimation and Testing of hypothesis.

Hypothesis: To decide whether to accept or reject a statement about the parameters. This statement is called hypothesis.

Statistical Hypothesis: A statistical hypothesis is a statement about the parameters of one or more populations.

ex 1) The majority of mean in the city are smokers.

2) The teaching methods in both the schools are effective.

Null hypothesis:

A null hypothesis is a statistical hypothesis formulated for the sole purpose of rejecting (accepting) it.

It is denoted by H_0 set up $H_0: \mu = \mu_0$

Ex: If we wish to decide whether one procedure is better than the other, Then, we formulate the null hypothesis as there is no difference between the procedures.

Alternative hypothesis

An alternative hypothesis is any statistical hypothesis that differs from a given null hypothesis

it is denoted by H_1

set up $H_1 : \mu > \mu_0$ (Right tailed)
$H_1 : \mu < \mu_0$ (Left tailed). $\Big\}$ one tailed

$H_1 : \mu \neq \mu_0$ (two tailed)

Level of significance: α

Test statistic : z

conclusion:

A region corresponding to a statistic 'z', in the sample spaces which leads to the rejection of H_0 is called Critical Region (or) Rejection Region. Those region which lead to the acceptance of H_0 give us a region called Acceptance Region.

Type I error:

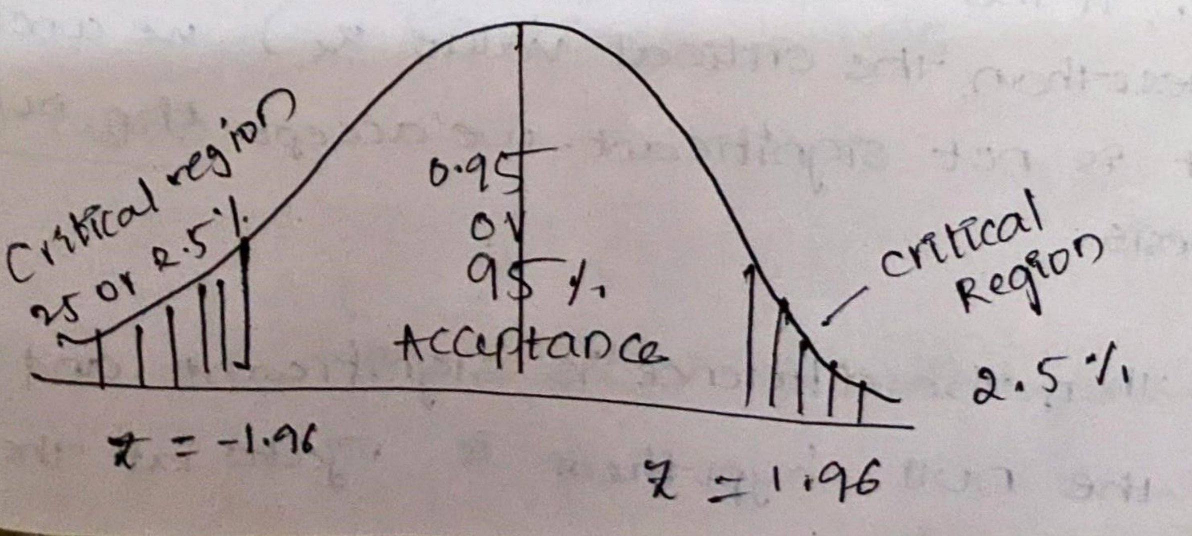
It is the error of rejecting Null Hypothesis (H_0) when it is true. The probability of making a type I error is denoted by "α", the level of significance and the probability of making correct decision is "$1-\alpha$".

Type II error:

It is the error of accepting the null hypothesis (H_0) when it is false. The probability of making a type II error is denoted by "β".

hypothesis - I

critical Region:



Two Tailed Test	$\lvert z_{\frac{\alpha}{2}} \rvert = 2.58$		
Right tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left tailed test	$-z_\alpha = -2.33$	$-z_\alpha = -1.645$	$-z_\alpha = -1.28$

Procedure for Testing of Hypothesis:

Null hypothesis; Set up the null hypothesis $H_0: \mu = \mu_0$

Altanative hypothesis: set up the Alternative hypothesis

$H_a: \mu > \mu_0$ (Right tailed)

$H_a: \mu < \mu_0$ (Left tailed)

$H_a: \mu \neq \mu_0$ (Two tailed)

Level of significance: α

Test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ where $\bar{x}$ = mean of samp

μ = mean of population

n = sample size

σ = standrad deviation

Conclusion: we compare the computed value of the test statistic z with the critical value z_α at given level of significance (α).

If $|z| < z_\alpha$ (i.e., if the absolute value of the calculated of z is less than the critical value z_α) we conclude that it is not significant. we accept the null hypothesis.

If $|z| > z_\alpha$ then the difference is significant and hence the null hypothesis is rejected at the level of significance

29/4/24

Note: when the size of the size of the sample(n) is greater than 30, then the sample is called large sample.

$n \geq 30$.

problem

It is claimed that a random sample of 49 tyres has a life of 15200 km. This sample was drawn from a population whose mean is 15150 km and a standrad deviation of 1200 km. Test the significance at 0.05 level.

Sol^n sample size (n) = 49

$\mu = 15150$

$\bar{x} = 15200$

$\sigma = 1200$

null hypothesis : $H_0 : \mu = 15150$

Alternative hypothesis : $H_1 : \mu \neq 15150$

Level of significance $(\alpha) = 0.05$

Test statistic : $Z_{cal} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{15200 - 15150}{1200/\sqrt{49}}$

$= \dfrac{50}{\frac{1200}{7}} \quad \dfrac{7}{}\ \dfrac{\cancel{350}}{\cancel{1200}}_{24} = \dfrac{7}{24} = 0.2917$

Z_α (table value) for 0.05 level of significance = 1.96

Conclusion :

$|Z|_{cal} < Z_{tab}$ we accept null hypothesis H_0

Q) An oceanographer wants to [illegible] depth of the ocean in a certain region is had previously been recorded 57.4 fathoms, as what can be concluded at the level of significance $\alpha = 0.05$, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standrad deviation of 5.2 fathoms.

size of sample $(n) = 40$, $\mu = 57.4$ $\bar{x} = 59.1$

$\sigma = 5.2$

Null hypothesis : $H_0 : \mu = 57.4$

Alternative hypothesis : $H_1 : \mu \neq 57.4$

Level of significance (α) : 0.05

Test statistic : $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{59.1 - 57.4}{\frac{5.2}{\sqrt{40}}}$

$= 2.06$

Z_{α} (tab) for 0.05 level of significance (two tailed test) is 1.96

Conclusion : $|Z|_{cal} > Z_{tab}$

The null hypothesis H_0 is Rejected.

Q) According to the norms established for a mechanical aptitude test, persons who are 18 years old have an average height of 73.2 with a standrad deviation 8.6. If 45 randomly selected persons of that age averaged 76.7, test the null hypothesis $\mu = 73.2$ aganist the alternative hypothesis

$\mu > 73.2$ at the 0.01 level of significance

Size of sample $(n) = 45$, $\mu = 73.2$ $\bar{x} = 76.7$
$\sigma = 8.6$

null hypothesis : $H_0 : \mu = 73.2$

Alternative hypothesis : $H_1 : \mu > 73.2$ (one tailed Right)

Level of significance (α) : 0.01

Test statistic : $Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{76.7 - 73.2}{8.6/\sqrt{45}}$

$= 2.73$

Z_α (tab) for 0.01 level of significance (Right tailed test) is 2.33

conclusion: $|Z|_{cal} > Z_{tab}$

Null hypothesis H_0 is Rejected (or) Reject N.H H_0

a) An ambulance service claims that it takes on the average less than 10 min. to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 min. and variance of 16 min. Test the claim at 0.05 level of significance.

Size of sample $(n) = 36$, $\mu = 10$, $\bar{x} = 11$, $\sigma = 4$

null hypothesis : $H_0 : \mu = 10$

Alternative hypothesis : $H_1 : \mu < 10$ (one tailed ~~Right~~ Left tailed)

Level of significance (α) : 0.05

Test statistic : $Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 10}{4/\sqrt{36}} = 1.5$

Z_α(tab) for 0.05 level of significance (left tailed test) is -1.645

Conclusion: $|Z|_{cal} > Z_{tab}$

reject H_0

H/W

2/5/24

2) A sample of 400 items is taken from a population whose standrad deviation is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38.

Test of significance for difference of two means (or) inferences concerning two means

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$ or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$

level of significance : α

Test statistic : $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ $\bar{x}_1, \bar{x}_2$ means of sample

σ_1^2, σ_2^2 variances, n_1 n_2 sizes of samples

conclusion : Accept / Reject H_0

1) A simple sample of heights of 6,400 Englishmen has a mean of 67.85 inches and S.D. 2.56 inches, while a simple sample of heights of 1,600 Australians has a mean of 68.55 inches and a S.D. of 2.52 inches. Do the data indicate the Australians are on the average, taller than Englishmen?

Given $n_1 = 6400$, $n_2 = 1600$

$\bar{x}_1 = 67.85$ $\sigma_1 = 2.56$ $\sigma_1^2 = (2.56)^2 =$

$\bar{x}_2 = 68.55$ $\sigma_2 = 2.52$ $\sigma_2^2 = (2.52)^2$.

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 < \mu_2$

Level of significance $(\alpha) = 0.05$

Test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{67.85 - 68.55}{\sqrt{\dfrac{6.5536}{6400} + \dfrac{6.304}{1600}}}$

$= -9.906$

$|z|_{cal} = 9.906$

z_{tab} value with 0.05 level of significance

$|Z|_{cal} > Z_{tab}$

$\therefore$ null hypothesis H_0 is rejected

2) Samples of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and shown below. make a large sample test to the significance of the difference b/w the means.

	Mean	S.D	size of sample
University A	55	10	400
University B	57	15	100

Sol^n Given

$n_1 = 400$, $n_2 = 100$

$\bar{x}_1 = 55$, $\bar{x}_2 = 57$

$\sigma_1 = 10$, $\sigma_2 = 15$

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$

Level of significance (α): 0.05

Test statistic $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{55-57}{\sqrt{\dfrac{(10)^2}{400} + \dfrac{(15)^2}{100}}} = -1.21$

$|Z|_{cal} = 1.26$

Z_{tab} with level of sign 0.05 is $= 1.996$

$|Z|_{cal} < Z_{tab}$

$\therefore$ null hypothesis H_0 accepted (or) Accept N.H H_0

3) A researches wants to know the intelligence of students in a school. he selected two groups of students. in the first group there 150 stud-ents having mean IQ of 75 with S.D of 15 in the second group there are 250 students having mean IQ of 70 with S.D of 20.

Sol^n Given

$n_1 = 150 \quad n_2 = 250$

$\bar{x}_1 = 75 \quad \bar{x}_2 = 70$

$\sigma_1 = 15 \quad \sigma_2 = 20$

null hypothesis $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$

level of significance $\alpha = 0.05$

Test Statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \dfrac{75 - 70}{\sqrt{\frac{30}{150} + \frac{40}{250}}} = 1.09$

$|z|_{cal} = 1.09$

z_{tab} with level of sign 0.05 is $= 1.96$

$|z|_{cal} < |z|_{tab}$ $\therefore$ null hypothesis accepted

4) A company claims that is its light bulbs are superior to those of its main competion. If a study showed that a sample of $n_1 = 40$ of its bulb that a mean life time of 647 hours of continuous use with a S.D of 27 hours; while a sample of $n_2 = 40$ bulbs made by its main com-petitor had a mean life time of 638 hours of continous use with a S.D of 31 hours. Does this substantiate the claim at the 0.05 level of significance. Given

$n_1 = 40 \quad , n_2 = 40$

$\bar{x}_1 = 647 \quad \bar{x}_2 = 638$

$\sigma_1 = 27$, $\sigma_2 = 31$

null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Level of significance (α) : 0.05

Test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{647 - 638}{\sqrt{\dfrac{(27)^2}{40} + \dfrac{(31)^2}{90}}} = 1.38$

$|z|_{cal} = 1.38$

z_{tab} with level of significance 0.05 = 1.96

$|z|_{cal} < z_{tab}$

$\therefore$ null hypothesis H_0 Accepted.

5) Suppose that we want to investigate whether on the average man earn morethan 20 per week morethan women in a certain industry. If sample data show that 60 men earn on the average $\bar{x}_1 = 292.50$ per week with a S.D of 15.6 while 60 women earn on average $\bar{x}_2 = 266.10$ per week was S.D of 18.20. What can we conclude at the 0.01 level of significance.

Given,

$n_1 = 60$, $n_2 = 60$

$\bar{x}_1 = 292.50$, $\bar{x}_2 = 266.10$

$\sigma_1 = 15.6$, $\sigma_2 = 18.20$

Null hypothesis : $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Level of significance (α) = 0.01

Test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{292.50 - 266.10}{\sqrt{\dfrac{(15.6)^2}{} + \dfrac{(18.20)^2}{}}}$

Z_{tab} with level of significance 0.01 = 2.56

$|z|_{cal}$ z_{tab}

3/5/24

Hypothesis - II

Testing of hypothesis for single proportions

Null Hypothesis H_0: $P = P_0$

Alternative hypothesis : H_1 : $P \neq P_0$ or $P < P_0$ or

$P > P_0$

Level of significance : α (0.01, 0.05)

Test statistic : $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$ $p \rightarrow$ sample proportion

numbers of success $p = \dfrac{x}{n}$

n = sample size

Conclusion : Accept | Reject H_0

1) Experience had shown that 20% of a manufactured product is of the top quality. In one day's production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level.

Sol^n $n = 400$

$P = 20\% = 0.2$ $P + Q = 1$

$Q = 1 - P = 0.8$ $Q = 1 - P$

$p = \dfrac{x}{n} = \dfrac{50}{400} = 0.125$ $p + q = 1$

$q = 1 - p = 0.875$

Null hypothesis H_0 : $P = P_0$

Alternative hypothesis H_1 : $P \neq P_0$

Level of significance $\alpha = 0.05$

Test statistic $(z) = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125-0.2}{\sqrt{\frac{(0.2)(0.8)}{400}}} = -3.75$

$|z|_{cal} = 3.75$

z_{tab} with 0.05 level of significance 1.96

$|z|_{cal} > z_{tab}$

Reject N.H H_0

(or)

null hypothesis H_0 is Rejected.

2) in a study design to investigate whether Certain detonators used with explosives in coal mining meet the requirement that at least 90%. will ignite the explosive when charged, it is found that 174 of 200 detonators function properly. Test the null hypothesis $p = 0.9$ against the alternative hypothesis $P < 0.9$ at the 0.05 level of significance

Sol^n:

$n = 200$

$P = 0.9$

$Q = 1-P = 0.1$

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

$q = 1-p = 0.13$

Null hypothesis H_0: $P = P_0$

Alternative hypothesis H_1: $P < P_0$ (Left Tailed)

Level of significance (α): 0.05

Test Statistic $(z) = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.87-0.9}{\sqrt{\frac{(0.9)(0.1)}{200}}} = -1.414$

$|Z|_{cal} = 1.414$

Z_{tab} with 0.05 level of significance 1.645

$|Z|_{cal} < Z_{tab}$

Accept N.H H_0

Q) A random sample of 1200 apples was taken from a large consignment and found that 10% of them are bad. The supplier claims that only 2% are bad. Test his claim at 95% level.

Soln

$n = 1200$

$P = 2\%$

$Q = 1 - P = \cancel{0.08}$ 0.98

$\cancel{x = \frac{10}{1200} \times 100 = 8}$

$x = \frac{10}{100} \times 1200 = 120$

$p = \frac{x}{n} = \frac{\cancel{120}}{\cancel{1200}}$ (1/10)

$= 0.1 \quad q = 1 - p = 0.9$

null hypothesis : H_0 : $P = P_0$

Alternative hypothesis H_1 : $P \neq P_0$

level of significance (α) : 0.05

$$\text{Test statistic } (Z) = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.1 - 0.02}{\sqrt{\frac{(0.02)(0.98)}{1200}}}$$

$|Z|_{cal} = 19.79$

Z_{tab} with 0.05 level of significance 1.96

$\therefore \quad |Z|_{cal} > Z_{tab}$

Rejected

Q) A manufacturer of electric bulbs claims that the percentage defectives in his product does not exceed 6. A sample of 40 bulbs is found to contain 5 defectives. Would you consider the claim justified?

Solⁿ: $n = 40$

$P = 6\% = 0.06$

$Q = 1-P = 0.94$

$p = \frac{x}{n} = \frac{5}{40} = 0.125$

$q = 1-p =$

Null hypothesis $H_0: P = P_0$

Alternative hypothesis $H_1: P < P_0$ (L.T.T)

Level of significance (α) = 0.05

Test Statistic (z) = $\frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125 - 0.06}{\sqrt{\frac{(0.06)(0.94)}{40}}} = 1.73$

$|z|_{cal} = 1.73$

z_{tab} with 0.05 level of significance 1.96

$|z|_{cal} < z_{tab}$

$\Rightarrow$ Accept n.H H_0 //

Q) 35 people were attacked by a disease and only 18 survived. will you reject the hypotheses that the survival rate if attacked by the disease is 85% in favour of the hypothesis that is more at 5% level.

Solⁿ $n = 35$

$P = 85\% = 0.85$

$Q = 1-P = 0.15$

$p = \frac{x}{n} = \frac{18}{35} = 0.51$

$q = 1-p =$

Null hypothesis $H_0: P = P_0$

Alternative hypothesis H_1: ~~$P<P_0$~~ $P > P_0$ (Right Tailed

Level of significance (α) = 0.05

Test statistic (z) = $\dfrac{p-P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.51-0.85}{\sqrt{\frac{(0.85)(0.15)}{35}}} = -5.63$

$|z|_{cal} = 5.63$

z_{tab} with 0.05 level of significance = 1.645 (R.T)

$|z|_{cal} > z_{tab}$

$\therefore$ Reject H_0

Q) In a large consignment of oranges, a random sample of 64 oranges revealed that 14 oranges were bad. Is it reasonable to ensure that 20% of the oranges are bad?

Sol

$n = 64$

$P = 20\% = 0.2$

$Q = 1-P = 0.8$

$p = \dfrac{x}{n} = \dfrac{14}{64} = 0.21$

Null hypothesis H_0 : $P = P_0$

Alternative hypothesis H_1 : $P \neq P_0$

Level of significance (α) : 0.05

Test statistic (z) = $\dfrac{p-P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.21-0.2}{\sqrt{\frac{(0.2)(0.8)}{64}}} = 0.2$

$|z|_{cal} = 0.2$

z_{tab} with 0.05 level of significance = 1.96 (two tailed)

$|z|_{cal} < z_{tab}$

$\therefore$ Accept N.H H_0

Q) In a random sample of 125 said they prefer thumps up to pepsi. Test the null hypothesis $p = 0.5$ against the alternative hypothesis $p > 0.5$.

Sol^n

$n = 125$

$P = 0.5$

$Q = 0.5$

$p = \frac{68}{125} = 0.54$

null hypothesis H_0: $P = P_0$

Alternative hypothesis H_1: ~~$P < P_0$~~ $P > P_0$ (Right T)

Level of significance (α) : 0.05

Test statistic: $\frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{(0.54)(0.5)}{125}}} = 0.894$

$|z|_{cal} = 0.899$

Z_{tab} at 0.05 level of significance is 1.645

$|z|_{cal} < Z_{tab}$

Accept null hypothesis H_0.

Test of significance for Difference proportions

null hypothesis H_0 : $P_1 = P_2$

Alternative hypothesis H_1 : $P_1 \neq P_2$

Level of significance (α) : $(0.01, 0.05)$

Test statistic $Z = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

where $P_1 = \frac{x_1}{n_1}$ $P_2 = \frac{x_2}{n_2}$ $\hat{p} = \frac{n_1P_1 + n_2P_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$

conclusion : Accept | Reject H_0

Note 1: suppose if the population proportions P_1 and P_2 are given to be distinctly different, then to test the difference in population proportions, the test statistic is

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

2. sample proportions are not given, the samply will not reveal the difference in the population proportions, the test statistic becomes

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Q) A manufacturer of electronic equipment subjects samples of two computing brands of transistors to an accelerated performance test. If 45 of 180 transistors of the first kind and 34 of 120 transistors of the second kind fail the test, what can be conclude at the level of significance $\alpha = 0.05$ about the difference b/w the corresponding sample proportions?

Soln Given $n_1 = 180$ $n_2 = 120$

$p_1 = \frac{x_1}{n_1} = \frac{45}{180}$ $p_2 = \frac{x_2}{n_2} = \frac{34}{120}$

$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{45 + 34}{180 + 120} = 0.263$

$\hat{q} = 1 - \hat{p} = 0.737$

null hypothesis H_0: $P_1 = P_2$

Alternative hypothesis H_1: $P_1 \neq P_2$

Level of significance $\alpha = 0.05$

Test statistic: $z = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.25 - \ldots}{\sqrt{(0.263)(0.737)\left(\frac{1}{400} \ldots\right.}}$

$|z|_{cal} = 0.636$

z_{tab} from 0.05 level of significance $= 1.96$

$|z|_{cal} < z_{tab}$

$\Rightarrow$ Accept N.H H_0 //

Q) In a city A, 20% of a random sample of 900 school boys had a certain slight physical defect, in another city B, 18.5% of a random sample of 1600 school boys had the same effect. Is the difference b/w the proportions significant at 0.05 level?

Solⁿ Given $n_1 = 900$, $n_2 = 1600$, $x_1 = \frac{20}{100} \times 900 = 180$

$x_2 = 296$ $\quad p_1 = \frac{x}{n_1} = \frac{180}{900} = 0.2$ $\qquad p_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$

$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} \Rightarrow \frac{180 + 296}{900 + 1600} = 0.1904$

$\hat{q} = 1 - \hat{p} = 1 - 0.19 = 0.8096$

Null hypothesis H_0: $P_1 = P_2$

Alternative hypothesis H_1: $P_1 \neq P_2$

Level of significance α: 0.05

Test statistic: $z = \frac{p_1 - p_2}{\sqrt{\hat{p}_1\hat{q}_1\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.2 - 0.185}{\sqrt{(0.1904)(0.8096)\left(\frac{1}{900}+\frac{1}{1600}\right)}}$

$z_{cal} = 0.916$

z_{tab} of 0.05 level of significance is 1.96

$|Z|_{cal} < Z_{tab}$

∴ Accept N.H H_0

Q) 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 from the two populations.

$n_1 = 1200$

$n_2 = 900$

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

$Q_1 = 0.7$ $\quad Q_2 = 0.75$

Null hypothesis H_0 : $P_1 = P_2$

Alternative hypothesis H_1 : $P_1 \neq P_2$

Level of significance (α) : 0.05

Test statistic $Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} = \frac{0.3 - 0.25}{\sqrt{\frac{(0.3)(0.7)}{1200} + \frac{(0.25)(0.75)}{900}}}$

$Z_{cal} = 2.55$

Z_{tab} of 0.05 level of significance is 1.96

$|Z|_{cal} > Z_{tab}$

∴ Rejected H_0

Q) A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8%. If it is found that 42 out of a sample 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim

27/5/24

Students t - test for difference of means.

null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ or

$\mu_1 < \mu_2$ or

$\mu_1 > \mu_2$

Level of significance : α (Degree of freedom $n_1 + n_2 - 2$)

Test statistic : $t_{cal} = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

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

or

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

Conclusion : Accept / Reject H_0

problem 1 :

Find the maximum difference that we can eapect with probability 0.95 b/w the mean of example of size 10 and 12 from a normal populaltion if their Standrad deviations are found to be

2 and 3 respectively.

Sol^n we have size of the first sample $n_1 = 10$

we have size of the second sample $n_2 = 12$

Standard deviation of first sample $s_1 = 2$

standard deviation of second sample $s_2 = 3$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{10 \times 2^2 + 12 \times (3)^2}{10 + 12 - 2}$$

$s^2 = 7.4$

$s = \sqrt{7.4} = 2.72$

① Null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Level of significance $(\alpha) = 0.05$

Test statistic $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$|\bar{x} - \bar{y}| = |t| \times s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Tabulated value of 't' for

$n_1 + n_2 - 2$

$10 + 12 - 2 = 20$

diff at 5%.

$(2.086)(2.72)\sqrt{\frac{1}{10} + \frac{1}{12}}$

$= 2.429$ //

2) Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	...

Test whether the two horses have the same running capacity

Sol^n Given,

size of the first sample $n_1 = 7$

size of the second sample $n_2 = 6$

$\bar{x}$ = mean of 1st sample

$$= \frac{28+30+32+33+33+29+34}{7} = 31.28$$

$\bar{y}$ = Mean of 2d sample

$$\frac{29+30+30+24+27+29}{6} = 28.16$$

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

Table

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
28	-3.28	10.75	29	0.84	0.705
30	-1.28	1.638	30	1.84	3.385
32	0.72	0.518	30	1.84	3.385
33	1.72	2.958	24	-4.16	17.305
33	1.72	2.958	27	-1.16	1.345
29	-2.28	5.198	29	0.84	0.705
34	2.72	7.398			

$\Sigma(x-\bar{x})^2$ $\quad \Sigma(y-\bar{y})^2$

$= 51.32 \quad = 26.83$

$$s^2 = \frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2} = 5.28$$

null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

Level of significance $(\alpha) : 0.05$

Test statistics $t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 2.44$

Tabulated value of t for $7+6-2 = 11$

$t_{cal} = 2.4, \; t_{tab} = 2.2$

$|t_{cal}| > t_{tab}$

$\therefore$ reject H_0

30/5/24

F-Test

To test whether is any significant difference b/w two estimates of population variance

To test the two samples have came from the same population, we use F-Test

null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

Alternative hypothesis : $H_1 : \sigma_1^2 \neq \sigma_2^2$

Level of significance : α (Degree of freedom $\nu_1 : n_1 - 1$, $\nu_2 : n_2 - 1$)

Test statistics : $F = \frac{\text{Greater variance}}{\text{Smaller variance}}$

where $s_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1}$, $s_2 = \frac{\Sigma(y_i - \bar{y})}{n_2 - 2}$

note: if sample variance s^2 is given, we can obtain population variance s^2 by using the relation $ns^2 = (n-1)S^2$

Conclusion: Accept / Reject H_0

Q1) The measurements of the output of two units have given the following results. Assuming that both samples have been obtained from the normal population at 5% significant level, test whether the two populations have the same variance.

unit - A	14.1	10.1	14.7	13.7	14.0
unit - B	14.0	14.5	13.7	12.7	14.1

Sol^n

null hypothesis : $\sigma_1^2 = \sigma_2^2$

Alternative hypothesis H_1 : $\sigma_1^2 \neq \sigma_2^2$

Level of significance α : 5% : 0.05

Test statistics $\nu_1 = n_1 - 1 = 4$

$\nu_2 = n_2 - 1 = 4$

$F_{(4,4)}(0.05) = 6.39$

A : $\bar{x} = \frac{14.1 + 10.1 + 14.7 + 13.7 + 14.0}{5} = 13.3$

B: $\bar{y} = \dfrac{14.0 + 14.5 + 13.7 + 12.7 + 14.1}{5} = 13.8$

Test statistic:

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

$= \dfrac{1}{4} \left[(14.1 - 13.3)^2 + (10.1 - 13.3)^2 + (14.7 - 13.3)^2 + (13.7 - 13.3)^2 + (14.0 - 13.3)^2 \right] = 3.37$

$S_2^2 = \dfrac{1}{4} \left[(14.0 - 13.8)^2 + (14.5 - 13.8)^2 + (13.7 - 13.8)^2 + (12.7 - 13.8)^2 + (14.1 - 13.8)^2 \right]$

$= \dfrac{1}{4} [\ ] = 0.46$

$\therefore F = \dfrac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$

$= \dfrac{3.37}{0.46} = 7.32$

since $F_{cal} > F_{tab}$, reject H_0

2) The time taken by workers in performing a Job by method I and method II is given below

method I	20	16	26	27	23	22	--
method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

Q) To examine the hypothesis
are more intelligent than the wives, a[...]
-ator took a sample of 10 couples and admi[...]
-tered them a test which measures the I.Q
The results are as follows.

husbands	117	105	97	105	123	109	86	78	103
wives	106	98	87	104	116	95	90	69	108

Test the hypothesis with a reasonable test at the level of significance of 0.05.

Sol^n we have $n_1 = 10$

$n_2 = 10$

$$\bar{x} = \frac{117+105+97+105+123+109+86+78+103+10}{10}$$

$= 103$

$$\bar{y} = \frac{106+98+87+104+116+95+90+69+108+85}{10}$$

$= 95.8$

Table:

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
117	14	196	106	10.2	104.04
105	2	4	98	2.2	4.84
97	-6	36	87	-8.8	77.44
105	2	4	104	8.2	67.24
123	20	400	116	20.2	408.04
109	6	36	95	-0.8	0.64
86	-17	289	90	-5.8	33.64
78	-25	625	69	-26.8	718.24
3	0	0	108	12.2	148.84
7	4	16	85	-10.8	

$\Sigma(x-\bar{x})^2$ $\qquad$ $\Sigma(y-\bar{y})^2$

$= 1606$ $\qquad$ $= 1679.6$

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

$$\frac{1606 + 1679.6}{10+10-2} = \frac{3285.6}{18} = 182.53$$

$s = 13.51$

1. Null hypothesis: $H_0 : \mu_1 = \mu_2$ (i.e no diff in IQ)
2. Alternative hypothesis : $\mu_1 > \mu_2$ (i.e, husbands are more intelligent then wives one tailed right)
3. Level of significance α: 0.05
4. Test statistic: $t = \dfrac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{103-95.8}{(13.51)\sqrt{\frac{1}{10}+\frac{1}{10}}}$

$\boxed{t_{cal} = 1.1916}$, $t_{tab} = 1.734$

$t_{cal} < t_{tab}$

Conclusion: $\therefore$ Accept H_0

Q) The soldiers participated in a shooting competition in the first week. After intensive training they participated in the competition in the second week. Their scores before and after training are given as follows.

Scores before	67	24	57	55	69	54	56	65	33	43
Score after	70	38	58	58	56	67	68	75	42	38

Do the data indicate that the soldiers have benefitted by the training?

Paired t - test

Two samples are said to be dependent when the elements in one sample are related to those to the other in any significant manner in fact two samples may consist of pairs of deviations made on some object, individual. The t-test is based on Paired observation. The paired observation can be calculated using.

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$ where $\bar{d}$ = Mean of difference

s = standard deviation

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