LARGE SAMPLE TEST

Vishwas Patil

Introduction:- Collecting information for statistical analysis is called collection of data. The aggregate of the objects of study is called population or universe.

There are two methods of collecting data :-

① **Census Method :-** Information is collected from every member of population.

② **Sampling Method :-** Information is collected from every member of a group collected from universe.

Vishwas Patil

Parameter and statistics :-

Mean, S.D. calculated from whole universe is called Parameter

Mean, S.D. calculated from sample is called statistic.

Note :- ① Let $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_n$ be the means of 'n' samples taken from the population with mean 'μ' then it can be proved that mean of means is always equal to 'μ'

$$\mu = (\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_n)/n$$

② **What is S.D. of sample means ? :-** With usual notation we can prove that S.D. of sample means (called standard error) is always equal to $\sigma/\sqrt{n}$ where σ is S.D. of the population and n is sample size.

Thus $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ and $V(\bar{x}) = \sigma^2/n$

then S.N.V. $z = (x-m)/\sigma$ reduces to $z = (\bar{x}-\mu)/\sigma/\sqrt{n}$

Procedure of Testing the Hypothesis :-

1) **Setting up Hypothesis :-** Hypothesis is a statement supposed to be true till it is proved false. It can be stated in various ways. For example parameter equal to given value, or greater than given value, or not equal to given value.

Here we have to state two Hypothesis instead of one. They are called as i) Null Hypothesis H_0 ii) Alternative Hyp. H_a they are set up in such a way that if one is true other is false.

3) Set up Level of significance (LOS) :- After setting up Null Hypothesis we set up the limits within which we expect the null hypothesis to lie. The limits are fixed depending upon the accurancy desired.

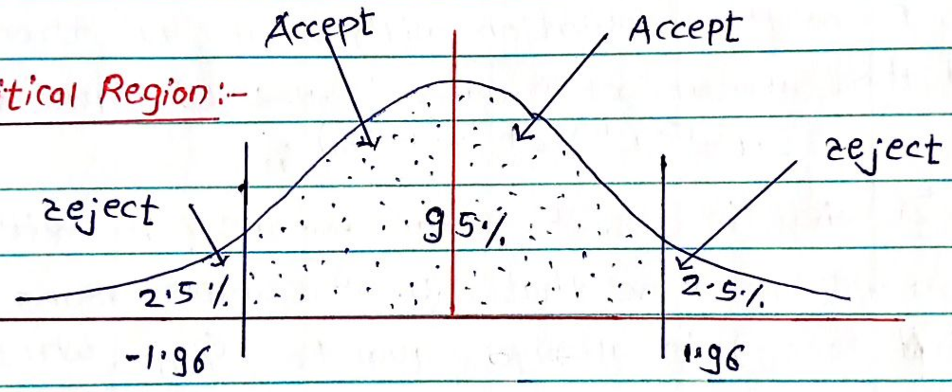
The probability that a random value of statistic will lie in the critical region is called the level of significance. $\alpha = 5\%$ LOS means the probability of rejecting a true hypothesis is 0.05

3) Test Statistics :- Depending uppon the nature of data and nature of problem we use Normal distribution, t-distribut^n, χ^2-distribution.

Vishwas Patil

LARGE - SAMPLE TEST ($n \geq 30$)

Note :- ① Critical Region :-



$P(-1.96 < z < 1.96) = 0.95$

Confidence level	Critical value of z
90%	1.64
95%	1.96
98%	2.33
99%	2.58

② Two tailed and one tailed Test :-

If the rejection area lies on two sides (two tails) the test is called two tailed otherwise it is one tailed test

i) In one tailed test hypothesis will be

Null Hypothesis $H_0 : \mu = \mu_0$

Alternate Hypothesis $H_a : \mu \neq \mu_0$

ii) ~~Rig~~ One tailed test

Null hypothesis $H_0 : \mu = \mu_0$

a) Right tailed test $H_a : \mu > \mu_0$

b) Left tailed test $H_a : \mu < \mu_0$

Table of Critical Value :-

	LOS			
Test	1%	2%	5%	10%
Two tailed	2·576	2·326	1·96	1·645
Right tailed	2·326	2·054	1·645	1·282
left tailed	-2·326	-2·054	-1·645	-1·282

Errors in Testing Hypothesis :-

We never know whether the hypothesis is true or false. Hence there arise four possibilities ;

i) True hypothesis is rejected ii) True hypothesis accepted

iii) false hypothesis is rejected iv) false hypothesis accepted

Type-I error :- This ~~hypothesis is~~ error arises when a true hypothesis is rejected i.e. when difference between sample value and hypothetical value exceeds the confidence limits. The error can be minimised by increasing limits, but then type-II error increases, because we never know whether the hypothesis is true or false in reality.

Type-II error :- This error arises when false hypothesis accepted, i.e. when the difference between sample value and hypothetical value lies within the limits. The error can be minimised by decreasing the limits but then type-II error increases.

Depending upon the nature of the problem one has to decide which error is serious and avoid it accordingly.

vishwasPatil

Type-I

Testing the hypothesis that population mean is μ :-

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

$H_a : \mu \neq \mu_0$ or $H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$

$$\boxed{z = (\bar{x} - \mu)/\sigma/\sqrt{n}}$$

Type-II

Testing the difference between means :- $\mu_1 - \mu_2$

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

Here null Hypothesis $H_0 : \mu_1 - \mu_2 = 0$ i.e. $\mu_1 = \mu_2$

or sometimes $\mu_1 - \mu_2 = a$ (given)

Alternate hypothesis $H_a : \mu_1 \neq \mu_2$ or $H_a : \mu_1 > \mu_2$ or $H_a : \mu_1 < \mu_2$

under $\mu_1 = \mu_2$ above formula reduces to

$$\boxed{z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}$$

Ex (1) A machine is claimed to produce nails of mean length 5 cm and S.D. 0.45 cm. A random sample of 100 nails gave 5.1 cm as their average length. Does the information of the machine justify the claim? Mention the level of significance you apply. ⟶ 2014, 2002, 14, 19, 19

Solⁿ:- given $\mu = 5$ cm $\sigma = 0.45$ cm , $n = 100$, $\bar{x} = 5.1$ cm.

Null Hypothesis $H_0 : \mu = 5$

Alternative Hypothesis $H_a : \mu \neq 5$ (two tailed)

$$z = (\bar{x} - \mu)/\sigma/\sqrt{n} = \frac{5.1 - 5}{0.45/\sqrt{100}}$$

$z = 2.22 \implies \boxed{|z| = 2.22}$ calculated value

Consider 5% LOS: $\implies$ table value of z at 5% LOS for two tailed test is $z_{0.05} = 1.96$

Calculated value > table value

$\implies$ Hypothesis is rejected

$\implies \mu \neq 5$

Ex-2 The mean height of random sample of 100 individuals from a population is 160. The S.D. of the sample is 10. Would it be reasonable to suppose that the mean height of population is 165?

⟶ 2005, 2011

Vishwas Patil

Soln:- Given $n=100$, $\bar{x}=160$, $\mu=165$, $\sigma=10$

Null Hypothesis $H_0: \mu=165$

Alternative Hypothesis $H_a: \mu \neq 165$ (two tailed)

$$z = \frac{\bar{x}_1-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{160-165}{10/\sqrt{100}} = -5$$

$\boxed{|z|=5}$ calculated value

consider 5% LOS: $\Longrightarrow$ table value of z at 5% LOS for two tailed test $z_{0.05}=1.96$

Calculated value > table value

$\Longrightarrow$ Hypothesis rejected

$\Rightarrow \mu \neq 165$

Ex 3 The average marks scored by 32 boys is 72 with S.D. 8 while that of 36 girls is 70 with S.D. 6. Test at 1% LOS whether the boys perform better than girls.

⟶ 2004, 2007, M-2014, D-2014, D-2015, 2016

Solution:- Given $n_1=32$, $\bar{x}_1=72$, $\sigma_1=8$, $n_2=36$, $\bar{x}_2=70$, $\sigma_2=6$

Null Hypothesis $H_0: \mu_1=\mu_2$ (boys not performing better than girls)

$H_a: \mu_1 > \mu_2$ (right tailed test)

$$z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} = \frac{72-70}{\sqrt{64/32+36/36}} = 1.15$$

$\Longrightarrow \boxed{|z|=1.15}$ calculated value

1% is LOS: $\Longrightarrow$ at 1% LOS table value of z for right tailed test is $z_{0.01}=2.326$

calculated value < table value

$\Longrightarrow$ Hypothesis accepted

boys do not perform better than girls.

Ex 4 Two samples drawn from two populations (different populations) gave the following results:

	Size	mean	S.D
Sample-I	125	380	25
Sample-II	150	340	30

test at 5% level of significance that the difference of the means of two populations is 45.

Solution:- Given $n_1 = 125$, $\bar{x}_1 = 340$, $\bar{x}_2 = 380$, $n_2 = 150$, $\sigma_1 = 25$, $\sigma_2 = 30$

Null Hypothesis $H_0: \mu_1 - \mu_2 = 45$

$H_a: \mu_1 - \mu_2 \neq 45$ (two tailed test)

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{340 - 340 - 45}{\sqrt{625/125 + 900/150}} = 1.5$$

$\Longrightarrow$ $\boxed{|z| = 1.5}$ calculated value

at 5% LOS : $\Longrightarrow$ table value of z at 5% LOS for two tailed test is $Z_{0.05} = 1.96$

Calculated value < table value

$\Longrightarrow$ Hypothesis is accepted.

$\Longrightarrow \mu_1 - \mu_2 = 45$

Vishwas Patil

Ex (5) The mean consumption of food grains among 400 sampled middle class cosumers is 380 grams per day per person with S.D. of 120 grams. A similar sample survey of 600 working class consumers gave a mean of 410 grams with S.D. of 80 grams. Are we justified in saying that the difference between the averages of two classes is 40? use 5% Level of significance.

Solⁿ:- Given $n_1 = 400$, $\bar{x}_1 = 380$, $\sigma_1 = 120$, $n_2 = 600$, $\bar{x}_2 = 410$, $\sigma_2 = 80$

Null Hypothesis $H_0: \mu_1 - \mu_2 = 40$

$H_a: \mu_1 - \mu_2 \neq 40$ (Two tailed test)

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{(380 - 410) - 40}{\sqrt{14400/400 + 6400/600}} = -10.24$$

$\Longrightarrow$ $\boxed{|z| = 10.24}$ Calculated value.

Given 5% LOS : $\Longrightarrow$ table value of z at 5% LOS for two tailed test is

$Z_{0.05} = 1.96$

calculated value > table value

$\Longrightarrow$ Hypothesis rejected

$\Longrightarrow \mu_1 - \mu_2 \neq 40$

Ex (6) The mean breaking strength of cables supplied by manufacturer is 1800 with S.D. 100. By new technique in the manufacturing process it is claimed that the breaking strength has increased. In order to test the claim a sample of 50 cables is tested. It is found that the mean breaking strength is 1850. Can we support the claim at 5·1% level of significance? ⟶ 2007, 2009, 2013

Solution:- Given $n=50$, $\bar{x}=1850$, $\mu=1800$, $\sigma=100$

Null Hypothesis $H_0: \mu=1800$

$H_a: \mu>1800$ (right tailed test)

Vighwas Patil

$$Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}=\frac{1850-1800}{100/\sqrt{50}}=3.54$$

$\Longrightarrow$ $\boxed{|Z|=3.54}$ calculated value

given 1·1% LOS: $\Longrightarrow$ table value of z at 1% LOS for right tailed test is $Z_{0.01}=2.326$

calculated value > table value

$\Longrightarrow$ Hypothesis is rejected

$\Longrightarrow$ $\mu>1800$ Breaking strength has increased.

Ex 7 The mean life of a sample of 100 electric light bulbs was found to be 1456 hours with S.D. 400. A second sample of 225 bulbs chosen from a different batch showed a mean life of 1400 hours with S.D. 144 hours. Assuming that the two populations, have same standard deviation find, if there any significant difference between the mean of two batches?

Solution:- Given $n_1=100$, $\bar{x}_1=1456$, $\sigma_1=400$

$n_2=225$, $\bar{x}_2=1400$, $\sigma_2=144$

Null Hypothesis $H_0: \mu_1=\mu_2$ (no significant difference)

$H_a: \mu_1\neq\mu_2$ (two tailed test)

$$Z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}=\frac{1456-1400}{\sqrt{(400)^2/100+(144)^2/225}}=1.84$$

$\Longrightarrow$ $\boxed{|Z|=1.84}$ calculated value

Consider 5% LOS: $\Longrightarrow$ at 5% LOS table value of z for two tailed test is $Z_{0.05}=1.96$

calculated value < table value

$\Longrightarrow$ Hypothesis accepted

$\Longrightarrow$ $\mu_1=\mu_2$ there is no significant difference.

Ex 8 A distribution with unknown mean μ has variance 1.5. Use central limit theorem to find how large a sample should be taken from the distribution in order that the probability will be at least 0.95 that the sample will be within 0.5 of the population mean?

Solution:- given variance $\sigma^2 = 1.5 \implies \sigma = \sqrt{1.5}$

sample will be within 0.5 of the population mean $\implies \bar{x} - \mu = 0.5$

for 0.95 probability value of $z = 1.96$ (0.95 prob. $\implies$ 5% LOS)

we have to find n:

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \implies 1.96 = \frac{0.5}{\sqrt{1.5}/\sqrt{n}}$$

Vishwas Patil

$$\implies (1.96)\sqrt{1.5} = \sqrt{n}\,(0.5)$$

$$\implies n = \frac{(1.96)^2(1.5)}{0.25} = 23.0496$$

$$\implies \boxed{n = 23}$$

SMALL SAMPLE TEST $(n<30)$

Student's t- distribution :-

[1] To test the hypothesis that population mean is μ.

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} \qquad \text{where } s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

[2] To test the hypothesis that two populations have same mean i.e. no significant difference between population means

If σ_1, σ_2 given then
$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}$$

for combined standard deviation $s (= \sigma_1 = \sigma_2)$

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{1/n_1 + 1/n_2}}}, \quad \text{where } s = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

we we are given S.D. of two samples

$$s_1 = \sqrt{\Sigma(x_1-\bar{x}_1)^2/n_1}, \quad s_2 = \sqrt{\Sigma(x_2-\bar{x}_2)^2/n_2}$$

then $$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}}$$

Note :- ① Degree's of freedom :- It is the number of values we are free to choose.

② Dependent Data :- In case of dependent data we first find the difference of the corresponding values of the two sets of data. obviously null Hypothesis $H_0 : \mu = 0$

$$\therefore \boxed{t = \frac{\bar{x}-0}{s/\sqrt{n-1}}} \qquad \text{where } s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

Ex ① Ten individuals are chosen at random from a population and their heights are found to be 63, 63, 64, 65, 66, 69, 69, 70, 70, 71 inches. Discuss the suggestion that mean height of universe is 65 inches. $\longrightarrow$ 2019

Soln :- Given $\mu = 65$, $n = 10$ $\qquad \bar{x} = \frac{\Sigma x}{n} = \frac{670}{10} \implies \bar{x} = 67$

Null Hypothesis $H_0 : \mu = 65$

$H_a : \mu \neq 65$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} \longrightarrow ① \qquad s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$x-\bar{x} =$ -4	-4	-3	-2	-1	2	2	3	3	4
$(x-\bar{x})^2 =$ 16	16	9	4	1	4	4	9	9	16

$\Sigma(x-\bar{x})^2 = 88 \Rightarrow S = \sqrt{\frac{88}{10}} \Rightarrow S = 3.13$

$① \Rightarrow t = \frac{67-65}{3.13/\sqrt{9}} \Rightarrow t = 1.91$

$\therefore \boxed{|t| = 1.91}$ calculated value

Consider 5% LOS: $\Rightarrow$ at 5% LOS table value of t for $10-1=9$ degree's of freedom is $t_{0.05} = 2.262$

calculated value < table value

$\Rightarrow$ Hypothesis accepted

$\therefore \mu = 65$ mean height of universe is 65 inches.

Ex ② A random sample of 16 observations has mean 103.75 cm. The sum of squares of the deviations from the mean is 843.75 cm. Can this sample be regarded as comming from the populations having 108.75 cm as the mean?

Soln:

Null Hypothesis $H_0: \mu = 108.75$

$H_a: \mu \neq 108.75$

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n-1}} \longrightarrow ①$$

given $n=16$, $\bar{x} = 103.75$, $\Sigma(x-\bar{x})^2 = 843.75$

$$\Rightarrow S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{843.75}{16}} = 7.26$$

$\therefore ① \Rightarrow t = (103.75 - 108.75)/(7.26/\sqrt{16}) = -2.67$

$\Rightarrow \boxed{|t| = 2.67}$ calculated value.

Consider 5% LOS: $\Rightarrow$ at 5% LOS table value of t for $16-1=15$ degree's of freedom $t_{0.05} = 2.131$

calculated value > table value

$\Rightarrow$ Hypothesis rejected.

i.e. $\mu \neq 108.75$

we can't say that mean of population is 108.75

Ex 3 If two independent random samples of sizes 15 and 8 have respectively the following means and population standard deviations,

$\bar{x}_1 = 980$, $\bar{x}_2 = 1012$, $\sigma_1 = 75$, $\sigma_2 = 80$

Test the hypothesis that $\mu_1 = \mu_2$ at 5% level of significance

(Assume the population to be normal)

Solution:- Given $\bar{x}_1 = 980$, $\bar{x}_2 = 1012$, $n_1 = 15$, $n_2 = 8$, $\sigma_1 = 75$, $\sigma_2 = 80$

Note:- Here we have assumed the population to be normal and population S.D. is given $\Longrightarrow$ use large sample test

Null Hypothesis $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$ (two tailed test)

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{980 - 1012}{\sqrt{(75)^2/15 + (80)^2/8}} = -0.93$$

$\Longrightarrow$ $\boxed{|Z| = 0.93}$ calculated value

Consider 5% LOS :- $\Longrightarrow$ table value of z at 5% LOS for two tailed test is $Z_{0.05} = 1.96$

calculated value < table value

$\Longrightarrow$ Hypothesis accepted.

Ex 4 Two independent samples of sizes 8 and 7 gave the following results:

Sample 1 : 19 17 15 21 16 18 16 14

Sample 2 : 15 14 15 19 15 18 16

Is the difference between sample means significant?

$\longrightarrow$ {2003, 04, 14}

Solution:- Null Hypothesis $H_0: \mu_1 = \mu_2$ (no significant difference)

$H_a: \mu_1 \neq \mu_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \longrightarrow (1) \quad \text{where} \quad S = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$n_1 = 8$, $n_2 = 7$ $\quad \bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{136}{8} \Longrightarrow \bar{x}_1 = 17$

$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{112}{7} \Longrightarrow \bar{x}_2 = 16$

$\Sigma(x_1 - \bar{x}_1)^2 = 4+0+4+16+1+1+1+9 \Rightarrow \Sigma(x_1 - \bar{x}_1)^2 = 36$

$\Sigma(x_2 - \bar{x}_2)^2 = 1+4+1+9+1+4+0 \Longrightarrow \Sigma(x_2 - \bar{x}_2)^2 = 20$

$\therefore S = \sqrt{\frac{36+20}{8+7-2}} \Longrightarrow S = 2.08$ using this in (1)

$$t = \frac{(17-16)}{2.08\sqrt{\frac{1}{8} + \frac{1}{7}}} = 0.93 \Longrightarrow \boxed{|t| = 0.93} \text{ calculated value}$$

consider 5% LOS :- $\Longrightarrow$ table value of t at 5% LOS for $8+7-2=13$ degree's of freedom is $t_{0.05}=2.16$

calculated value < table value

$\Longrightarrow$ Hypothesis accepted.

$\therefore \mu_1=\mu_2$ there is no significant difference.

Ex ⑤ The mean and S.D. of heights of 8 randomly chosen soldiers are 166.9 cms and 8.29 cms respectively. The corresponding values of 6 randomly chosen sailors are 170.3 cms and 8.5 cm respectively. Based on this data can we conclude that the soldiers, in general, are shorter than the sailors

$\rightarrow$ 2005,

Solution :- Null Hypothesis $H_0: \mu_1=\mu_2$ (they are not)

$H_a: \mu_1<\mu_2$

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \longrightarrow ① \quad \text{where } S=\sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

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

given $n_1=8$, $\bar{x}_1=166.9$, $s_1=8.29$

$n_2=6$, $\bar{x}_2=170.3$, $s_2=8.5$

Now $S=\sqrt{\frac{8(8.29)^2+6(8.5)^2}{8+6-2}}$

$S=9.05$

$\therefore ① \Longrightarrow t=\frac{166.9-170.3}{9.05\sqrt{1/8+1/6}}=-0.6956$

$\Longrightarrow \boxed{|t|=0.6956}$ calculated value.

consider 5% LOS :- $\Longrightarrow$ at 5% LOS table value of t for $8+6-2=12$ degree's of freedom is $t_{0.05}=2.179$

calculated value < table value

$\Longrightarrow$ Hypothesis accepted.

$\mu_1=\mu_2$, they are not.

Ex ⑥ Ten school boys were given a test in statistics and their scores were recorded. They were given a months special coaching and second test was given to them in the same subject at the end of the coaching period. Test if the marks given below give evidence to the fact that students are benefitted by coaching.

Marks in I^{st} Test: 70 68 56 75 80 90 68 75 56 58

Marks in II^{nd} Test: 68 70 52 73 75 78 80 92 54 55 —S19-17,

Soln: Note that data is dependent ⇒ difference of values is 'x'

Null Hypothesis $H_0: \mu = 0$

$H_a: \mu \neq 0$

$$t = \frac{\bar{x} - 0}{S/\sqrt{n-1}} \quad \text{where } S = \sqrt{\frac{(x-\bar{x})^2}{n}}$$

Test-I :	~~70~~	~~68~~	56	75	80	90	68	75	56	58
Test-II :	68	70	52	73	75	78	80	92	54	55
$x = I - II$:	2	-2	4	2	5	12	-12	-17	2	3

$$\bar{x} = \frac{\Sigma x}{n} = \frac{-1}{10} \Longrightarrow \boxed{\bar{x} = -0.1}$$

$\Sigma(x-\bar{x})^2 =$ ~~...~~ 41

$$\Longrightarrow \boxed{\Sigma(x-\bar{x})^2 = \cancel{648.9}}\ 642.9$$

or $S = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{643}{10} - (0.1)^2}$

$S = 8.02$

$$S = \sqrt{\frac{\cancel{648.9}\ 642.9}{10}} = \cancel{8.05}\ 8.02$$

using this in ① $t = \frac{-0.1}{\cancel{8.05}/\sqrt{9}} = -0.036 \Rightarrow \boxed{|t| = 0.036}$

consider 5% LOS :⟹ at 5% LOS table value of t for $10-1=9$ d.f.

is $t_{0.05} = 2.262$

calculated value < table value

⟹ Hypothesis is accepted.

Ex ⑦ A certain injection administered to 12 patients resulted in the following changes of blood pressure:

5, 8, 2, -1, 3, 0, 6, -2, 1, 5, 0, 4

Can it be concluded that the injection will be in general accompnied by increase in blood pressure?

→ 98, 99, 2003, 04, 11, 16

Soln:-

Null Hypothesis $H_0: \mu = 0$ (dependent data why?)

Alternate hypothesis $H_a: \mu \neq 0$

$$t = \frac{\bar{x} - 0}{S/\sqrt{n-1}} \rightarrow ① \quad \text{where } S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$n = 12$, $\Sigma x = 31$

$$\Longrightarrow \bar{x} = \frac{\Sigma x}{n} = \frac{31}{12} \Longrightarrow \boxed{\bar{x} = 2.58}$$

x	5	8	2	-1	3	0	6	-2	1	5	0	4
$(x-\bar{x})^2$	5.86	29.38	0.34	12.82	0.18	6.66	11.70	20.98	2.5	5.86	6.66	2.02

$$\Longrightarrow \Sigma(x-\bar{x})^2 = 104.96 \Rightarrow S = \sqrt{\frac{104.96}{12}} \Rightarrow \boxed{S = 2.95}$$

using this in ① $\quad t = \dfrac{2.58}{2.95/\sqrt{11}} = 2.9 \Longrightarrow \boxed{|t| = 2.9}$

consider 5% LOS :- at 5% LOS table value of t for $12-1=11$ degree's of freedom is $t_{0.05} = 2.201$

calculated value > table value $\Longrightarrow$ Hypothesis rejected $\mu \neq 0$

Ex ⑧ In a certain experiment ~~two~~ to compare two type of pig-Foods A and B, the following results of increasing weights were obtained.

Pig Number	1	2	3	4	5	6	7	8
Increase in Weight x kg by A	49	53	51	52	47	50	52	53
Increase in weight y kg by B	52	55	52	54	50	54	54	53

i) Assuming that the two sample of pigs are independent, can we conclude that food B is better than Food A.

ii) Examine the case if same set of 8 pigs were used in both the case.

$\longrightarrow$ { 2004, 2006, 2010

Soln :- ① Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternate hypothesis $H_a: \mu_1 < \mu_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightarrow ① \quad \text{where} \quad S = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$$

$n_1 = 8$, $n_2 = 8$, $\Sigma x_1 = 407$, $\bar{x}_1 = \dfrac{\Sigma x_1}{n_1} = \dfrac{407}{8} \Longrightarrow \boxed{\bar{x}_1 = 50.875}$

$\bar{x}_2 = \dfrac{\Sigma x_2}{n_2} = \dfrac{423}{8} \Longrightarrow \boxed{\bar{x}_2 = 52.875}$

$\Sigma(x_1-\bar{x}_1)^2 = 3.515 + 4.515 + 0.015 + 1.265 + 15.015 + 0.765 + 1.265 + 4.515 = \boxed{30.875}$

$\Sigma(x_2-\bar{x}_2)^2 = 0.765 + 4.515 + 0.765 + 1.265 + 8.265 + 1.265 + 1.265 + 0.015 = \boxed{\cancel{18.42}}\ 16.875$

$$\therefore S = \sqrt{\frac{30.875 + 16.875}{8+8-2}} = \sqrt{\frac{47.75}{14}} = 1.77 = 1.77$$

using this in ① $\Longrightarrow t = \dfrac{(50.875 - 52.875)}{1.77\sqrt{\frac{1}{8}+\frac{1}{8}}} = -2.17 \Longrightarrow \boxed{|t| = 2.17}$

consider 5% LOS :- $\Longrightarrow$ at 5% LOS table value of t for $16-2=14$ degree's of freedom is 2.145

calculated value is > table value $\Longrightarrow$ Hypothesis is rejected. Food B is better than A.

ii) Same set of pig is used $\Longrightarrow$ data is dependent.

Null Hypothesis $H_0 : \mu = 0$

$H_a : \mu \neq 0$

$$t = \frac{\bar{x} - 0}{S/\sqrt{n-1}} \longrightarrow (1) \qquad S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$x = A - B =$	-3	-2	-1	-1	-3	-4	-2	0
$(x-\bar{x})^2 =$	1	0	1	4	1	4	0	4

$\Sigma x = -16$, $n = 8$

$\Longrightarrow \boxed{\bar{x} = -2}$

$\Longrightarrow \Sigma(x-\bar{x})^2 = 15 \Longrightarrow S = \sqrt{15/8} \Longrightarrow \boxed{S = 1.36}$

$$t = \frac{-2}{1.36/\sqrt{7}} = -3.89 \Longrightarrow \boxed{|t| = 3.89}$$

Consider 5% LOS :- $\Longrightarrow$ table value of t at 5% LOS for $8-1=7$ degree's of freedom is $t_{0.05} = 1.90$

calculated value > table value $\Longrightarrow$ Hypothesis is rejected.

$\Longrightarrow$ Food B is better than A.

CHI - SQUARE Distribution

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

where O: observed frequency   E: Expected (calculated) frequency

Use :→ ① It is used to determine association between two or more attributes.

② It is used to test the goodness of fit.

Note:- For association between two attributes data will be given in rows and columns. In such case degree's of freedom will be $(r-1)(c-1)$ where r: no. of rows , c: number of columns.

Ex ① The following table gives the number of accidents in a city during a week. Find whether the accidents are uniformly distributed over a week.

Day :	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
Number of Accidents :	13	15	9	11	12	10	14	84

Solution:- ⟶ 2009, 2017

Null Hypothesis H_0: Accidents are equally distributed

H_a: Accidents do not

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} \longrightarrow ①$$

If accidents are equally distributed then on all days of week, there will be $\left(\frac{84}{7}\right) = 12$ accidents on every day. where prob. of day $= \frac{1}{7}$

$\Rightarrow$ Expected frequency $= \frac{84}{7} = 12 \Rightarrow E$.

O :	13	15	9	11	12	10	14
E :	12	12	12	12	12	12	12
$\frac{(O-E)^2}{E}$:	$\frac{1}{12}$	$\frac{9}{12}$	$\frac{9}{12}$	$\frac{1}{12}$	0	$\frac{4}{12}$	$\frac{4}{12}$

$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 28/12$

$\Rightarrow \chi^2 = 2.33$

Consider 5% LOS : $\Rightarrow$ at 5% LOS table value of χ^2 for $7-1=6$ degree's of freedom $\chi^2_{0.05} = 12.59$

Calculated value < table value

$\Rightarrow$ Hypothesis accepted i.e. accidents are uniformally distributed over a week.

Ex (2) Fit Binomial distribution for the data below and test the goodness of fit.

x:	0	1	2	3	4	5	6
f:	5	18	28	12	7	6	4

Solution:- To fit Binomial distribution :- $p(x) = {}^nC_x\, p^x q^{n-x}$

$n = 6$, $\Sigma f = 80 = N$, $\Sigma(xf) = 192 \implies \text{mean} = np = \dfrac{\Sigma(xf)}{\Sigma f}$

$\therefore\ 6p = 2.4 \implies p = 0.4 \implies q = 0.6$

$\therefore\ p(x) = {}^6C_x (0.4)^x (0.6)^{6-x}$

to find expected frequencies:- $(\Sigma f)\, p(x) = N p(x) = 80\ {}^6C_x (0.4)^x (0.6)^{6-x}$

If $x = 0$, $80\ {}^6C_0 (0.4)^0 (0.6)^6 = 3.73$

If $x = 1$, $80\ {}^6C_1 (0.4) (0.6)^5 = 14.93$

If $x = 2$, $80\ {}^6C_2 (0.4)^2 (0.6)^4 = 24.88$

If $x = 3$, $80\ {}^6C_3 (0.4)^3 (0.6)^3 = 22.12$ $\implies$ Expected frequencies E

If $x = 4$, $80\ {}^6C_4 (0.4)^4 (0.6)^2 = 11.06$

If $x = 5$, $80\ {}^6C_5 (0.4)^5 (0.6)^1 = 2.95$

If $x = 6$, $80\ {}^6C_6 (0.4)^6 (0.6)^0 = 0.33$

To test the goodness of fit :-

Null Hypothesis H_0 : fit is good

H_a : Fit is not good

$$\chi^2 = \sum \left[\frac{(O-E)^2}{E}\right] \longrightarrow (1)$$

O : Observed frequency (given)

E : Expected frequency (calculated)

O :	5	18	28	12	7	6	4
E :	3.73	14.93	24.88	22.12	11.06	2.95	0.33
$\frac{(O-E)^2}{E}$	0.43	0.64	0.39	4.63	1.50	3.20	40.81

$(1) \Rightarrow \chi^2 = 15.59$

Consider 5% LOS :$\Rightarrow$ at 5% LOS table value of χ^2 For $7-2=5$ degree's of freedom is $\chi^2_{0.05} = 12.592$ 11.07

calculated value > table value $\implies$ Hypothesis Rejected.

i.e. fit is not good.

Ex (3) Theory predicts that the proportion of beans in the four groups A, B, C, D should be 9 : 3 : 3 : 1. In an experiment among 1600 beans the number in the four groups were 882, 313, 287, 118. Does the experimental results support the theory?

$\longrightarrow$ 2001, 2006, 2013

Soln:- Null Hypothesis H_0: Suppose the proportion be 9 : 3 : 3 : 1

H_a: Proportion is not in above ratio

$$\chi^2 = \sum \frac{(O-E)^2}{E} \longrightarrow (1)$$

O: observed frequency (given)

E: Expected Frequency (calculated)

$N = 1600$, Probability of A i.e. $P(A) = \frac{9}{16}$, $P(B) = \frac{3}{16}$, $P(C) = \frac{3}{16}$, $P(D) = \frac{1}{16}$

O	E	$(O-E)^2/E$
882	$1600(\frac{9}{16}) = 900$	0.36
313	$1600(\frac{3}{16}) = 300$	0.56
287	$1600(\frac{3}{16}) = 300$	0.56
118	$1600(\frac{1}{16}) = 100$	3.24

$(1) \Longrightarrow \chi^2 = 4.72$

consider 5% LOS: $\Longrightarrow$ at 5% LOS table value of χ^2 for $4-1=3$ degree's of freedom is

$$\chi^2_{0.05} = 7.81$$

calculated value < table value $\Rightarrow$ Hypothesis accepted

$\Longrightarrow$ Given proportion is correct

Ex (4) Fit Poisson distribution and test the goodness of fit.

x:	0	1	2	3	4	5
f:	142	156	69	27	5	1

$\longrightarrow$ {2003}

Soln:- To fit Poisson Distribution :- $p(x) = \frac{e^{-m} m^x}{x!}$

$m = \frac{\Sigma(xf)}{\Sigma f} = \frac{400}{400} \Longrightarrow \boxed{m = 1}$

$p(x) = \frac{e^{-1} 1^x}{x!} \Longrightarrow \boxed{p(x) = \frac{e^{-1}}{x!}}$

$\Sigma f = N = 400$

to find frequencies :- $N p(x) = 400 \, e^{-1}/x!$

If $x = 0$, $NP(0) = 400 e^{-1} = 147.15$

If $x = 1$, $400 \frac{e^{-1}}{1!} = 147.15$

If $x = 2$, $400 \frac{e^{-1}}{2!} = 73.57$

If $x = 3$, $400 e^{-1}/3! = 24.53$

If $x = 4$, $400 e^{-1}/4! = 6.13$

If $x = 5$, $400 e^{-1}/5! = 1.23$

$\Longrightarrow$ Expected frequencies E

O	142	156	69	27	5	1
E	147.15	147.15	73.57	24.53	6.13	1.23
$\frac{(O-E)^2}{E}$	0.18	0.53	0.28	0.25	0.21	0.04

we have used two conditi[ons] Σf and Σfx

$\Rightarrow$ d.f. = 6 - 2 = 4

Null Hypothesis H_0: Fit is good

H_a: Fit is not good

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 1.49$$

at 5% LOS table value of χ^2 for $6-2= 4$ degree's of freedom = 9.49

calculated value < table value

$\longrightarrow$ Hypothesis accepted, fit is good.

Ex (5) In an experiment on immunisation of cattle from Tuberculosis the following results were obtained

	Affected	Not affected	Total
Inoculated	267	27	294
Not inoculated	757	155	912
Total	1024	182	1206

use χ^2 test to determine the efficacy of vaccine in preventing tuber-culosis. ⟶ 2003, 2015

Solution:- Null Hypothesis H_0: there is no association.

H_a: there is association.

$$\chi^2 = \sum \frac{(O-E)^2}{E} \longrightarrow ①$$

O: observed frequency (given)

E: Expected frequency (calculated)

O	E		$(O-E)^2/E$
267	$\frac{1024 \times 294}{1206} =$	249.63	1.21
27	$\frac{182 \times 294}{1206} =$	44.37	6.79
757	$\frac{1024 \times 912}{1206} =$	774.37	0.39
155	$\frac{182 \times 912}{1206} =$	137.63	2.19

① $\Rightarrow \chi^2 = 10.58$

at 5% LOS table value of χ^2 for $(2-1)(2-1) = 1$ degrees of freedom is $\chi^2_{0.05} = 3.84$

calculated value > table value

$\Rightarrow$ Hypothesis rejected.

vaccine is effective

Ex (06) A total of 3759 individuals were interviewed in a public opinion survey on a political proposal. Of them 1872 were men and the rest women. A total of 2257 individuals were in favour of proposal and 917 opposed it. A total of 243 men were undecided and 442 women were opposed to it. Do you justify or contradict the hypothesis that there is no association between sex and attitude at 5% Level of significance. ⟶ 2009, 2007, 14

Soln:-

Null Hypothesis H_0: There is no association betn sex and attitude

H_a: there is association

$$\chi^2 = \sum \frac{(O-E)^2}{E} \longrightarrow ①$$

O: observed frequency (given)

E: Expected frequency (calculated)

	favour	opposed	undecided	total
Men	1154	475	243	1872
women	1103	442	342	1887
total	2257	917	585	3759

O: observed	E: Expected Frequency	$(O-E)^2/E$
1154	$\frac{2257 \times 1872}{3759} = 1124$	0.80
475	$\frac{917 \times 1872}{3759} = 456.67$	0.74
243	$\frac{585 \times 1872}{3759} = 91.33$	8.02
1103	$\frac{2257 \times 1887}{3759} = 1133$	0.79
442	$\frac{917 \times 1887}{3759} = 460.32$	0.73
342	$\frac{585 \times 1887}{3759} = 293.67$	7.95

(1) $\Rightarrow x^2 = 19.03$

at 5% LOS table value of x^2 for $(2-1)(3-1)=2$ degree's of freedom is $x^2_{0.05} = 5.91$

calculated value > table value $\Longrightarrow$ Hypothesis rejected

$\therefore$ there is no association between sex and attitude

Ex (7) out of 800 persons 25% were literate and 300 had travelled beyond the limits of the district 40% of the literates were among those who had not travelled. Prepare 2x2 table and test at 5% LOS whether there is relation between travelling and literacy.

Soln: Null Hypothesis H_0: there is no relation

H_a: There is a relation

$$x^2 = \sum\left[\frac{(O-E)^2}{E}\right] \longrightarrow (1)$$

O: observed frequency

E: Expected frequency

Total: 800, No. of literate $= 800(0.25) = 200 \Longrightarrow$ iletrate $= 600$

300 out 800 had travelled ; 40% of literates not travelled $= 200(0.40) = 80$

	Literate	iliterate	total
travelled	120	180	300
not travelled	80	420	500
total	200	600	800

O	E	$(O-E)^2/E$
120	$\frac{200 \times 300}{800} = 75$	27
180	$\frac{600 \times 300}{800} = 225$	9
80	$\frac{200 \times 500}{800} = 125$	16.2
420	$\frac{600 \times 500}{800} = 375$	5.4

$x^2 = 57.6$

at 5% LOS for $(2-1)(2-1)=1$ degree's of freedom $x^2_{0.05} = 3.84$

calculated value > table value

$\Longrightarrow$ Hypothesis rejected

$\therefore$ there is a relation between travelling and literacy.

	Dead	Surviving
inoc.	2	10
not inoc	8	4

Ex 8 Two batches of 12 animals each are given test of inoculation. One batch was inoculated and other was not. The number of dead and surviving animals are given in the following table for both cases. Can the inoculation be regarded as effective against the disease at 5% level of significance.

(make Yate's correction) $\longrightarrow$ 2004, 2012

Solution: In 2x2 table degree's of freedom will be $(2-1)(2-1)=1$.

If one of the cell frequency is less than 5, we have to use pooling method. But this will result in χ^2 with zeros degree of freedom which is meaning less. In such case Yate in 1934 suggest to use $\chi^2 = \sum \frac{[|O-E|-0.5]^2}{E}$ this is the improved value of χ^2.

Null Hypothesis H_0: No association (not effective)

H_a: effective

	D	S	Tot.
Ino	2	10	12
Not In	8	4	12
Total	10	14	24

$$\chi^2 = \sum\left(\frac{(|O-E|-0.5)^2}{E}\right) \longrightarrow ①$$

O	E	$\|O-E\|-0.5$	$\frac{[\|O-E\|-0.5]^2}{E}$
2	$\frac{10\times12}{24}=5$	2.5	1.25
10	$\frac{14\times12}{24}=7$	2.5	0.89
8	$\frac{10\times12}{24}=5$	2.5	1.25
4	$\frac{14\times12}{24}=7$	2.5	0.89

$① \Longrightarrow \chi^2 = 4.29$

Calculated value.

table value of χ^2 at 5% LOS for $(2-1)(2-1)=1$ degree's of freedom is $\chi^2_{0.05} = 3.81$

calculated value > table value

$\Longrightarrow$ Hypothesis rejected i.e. inoculation is efective.

Ex 09 The figures given below are ⓐ observed frequencies ⓑ frequencies of normal distribution, having same mean, S.D. and total frequency as in ⓐ

ⓐ 1, 12, 66, 220, 495, 792, 924, 792, 495, 220, 66, 12, 1

ⓑ 2, 15, 66, 210, 484, 799, 943, 799, 484, 210, 66, 15, 2

Apply χ^2 test of goodness of fit.

$\longrightarrow$ 2004,

Null Hypothesis H_0: Fit is good

H_a: Fit is not good

$$\chi^2 = \sum \frac{(O-E)^2}{E} \longrightarrow (1)$$

Since the frequencies at the beginning and end are less than 5, we group them and then apply χ^2

O	E	$(O-E)^2/E$
13	17	0.94
66	66	0
220	210	0.48
495	484	0.25
792	799	0.06
924	943	0.38
792	799	0.06
495	484	0.25
220	210	0.48
66	66	0
13	17	0.94

(1) $\Rightarrow \chi^2 = 3.84$ calculated value

There are originally 13 classes, since they reduced to 11 by grouping them twice the degree's of freedom reduced by 2.

Further since the mean, S.D, total frequency of original data is used, three restrictions are introduced reducing degree's of freedom by 3.

$\Longrightarrow$ degree's of freedom = 13-2-3 = 8

at 5% Los table value of χ^2 for 8 degrees of freedom is $\chi^2_{0.05} = 15.51$

Calculated value < table value

$\Longrightarrow$ Hypothesis accepted. Fit is good.