Unit 5

18 June 2023 03:00 PM

Unit - 5

Universe or Population

1. Finite
2. Infinite

Collection of Individual / Members or set of all experimental Data

Sample

A finite subset of Universe → Sample size

Sample size: The number of individuals in A sample

Sampling

Sampling → Process of selecting a sample

"Parametrs & statistics"

Population		Sample
μ (mu)	→ Mean ←	$\bar{x}$ (X bar)
σ^2 (Sigma square)	→ Variance ←	S^2 (S square)
σ (Sigma)	→ Standard Deviation ←	S (S)
N	→ Size ←	n

Hypothesis / statistical Hypothesis

Required to make decisions about population on the Basis of sample information

Test of Hypothesis → Deciding A Hypothesis is to be Accepted or rejected

1: Null Hypothesis (H_0)	→ Tested Under the Assumption it is Be
2: Alternative Hypothesis (H_1)	→ Complementary of the null Hypothesis

Thus. (H_0): Accept then (H_1): Reject

(H_1): Reject then (H_0): Accept

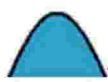
i.e $H_0 \cap H_1 = \phi$

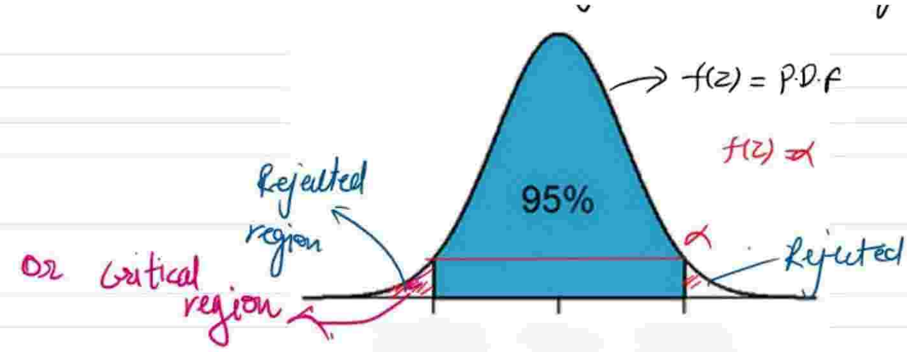
Test of Significance :-

Null Hypothesis Testing : The procedure / Test which enable us to whether the null Hypothesis is Accepted or Rejected on the basis of sample information

1. t- Test
2. F -Test
3. Z- Test
4. Chi- square Test (χ^2- Test)

Level of Significance :- The value of probability above which we not reject the null Hypothesis


→ $f(z) = P.D.F$



$\boxed{\text{Students } t\text{-Distribution or } t\text{-Test}}$:-

Note:- ① To test the Significance of the mean of a random sample.

② t-test is applicable when

(i) Sample size $n \le 30$

(ii) population S.D (σ) is unknown.

imp ③ To test the Significance difference between the sample mean and the hypothetical population mean.

Working Rule

① Null Hypothesis $\boxed{H_0 : \mu = \mu_0}$ (given)

"There is no Significant difference b/w sample mean and Hypothetical population Mean."

② Alternative Hypothesis : $\boxed{H_1 : \mu \neq \mu_0}$

③ T-test

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

"if S.D. of sample mean is not given then use it"

Here,

$\bar{x}$ = sample mean, n = Sample Size

s = S.D. of Sample

μ = population Mean.

If S.D of sample mean given "s" then $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

Also find $|t|$.

④. From the t-test table find the value of

t_α at level of significance

α $\boxed{5\% = \frac{5}{100} = 0.05 \text{ most cases}}$

for degree of freedom $\boxed{\text{"}n-1\text{"}}$

i.e; Find $t_{0.05}$

⑤ conclusion :-

(i) if $|t| < t_\alpha$ the H_0 is accepted

(ii) if $|t| > t_\alpha$ then H_0 is rejected.

Note:- ① If H_0 is accepted

⇒ there is no significance difference b/w sample mean and hypothetical population Mean.

⇒ Sample have been taken from the correct population

② If H_0 is rejected:-

⇒ There is significance difference b/w sample mean and hypothetical population mean.

Sample could not have come from this population

Fiducial limits of population Mean :-

$$\left|\frac{\bar{x}-\mu}{S/\sqrt{n}}\right| < t_\alpha \quad \text{for acceptance of } H_0$$

$$x - t_\alpha\, S/\sqrt{n} < \mu < \bar{x} + t_\alpha\, S/\sqrt{n}$$

95% Confidence limit (level of significance 5%) are $\bar{x} \pm t_{0.05} S/\sqrt{n}$

99% " " (" " " 1%) are $\bar{x} \pm t_{0.01} S/\sqrt{n}$

Q.1. Find the student's t value for following variable value in a sample of eight

−4, −2, −2, 0, 2, 2, 3, 3

Taking the mean of the universe (population) to be zero.

$-4, -2, -2, 0, 2, 2, 3, 3$

Taking the mean of the universe (population) to be zero.

Q. A machine which produce mica insulating washers of use in electric devices is set to turn out washers having a thickness of 10 miles (1 mil = 0.001 inch). A sample of 10 washers has an averge thickness of 9.52 mils. with a standard deviation of 0.60 mil. find out t.

Q. A sample of 20 items has mean 42 units and S.D 5 unit. Test the hypothesis it is a random sample from a normal population with mean 45 unit.

Q. A random sample of size have 16 and 53 as mean the sum of the deviation from mean is 135. can this sample be regarded as taken from the population having 56 as mean. obtain 95% and 99% confidence limits of mean of population

Q. The lifetime of electric bulbs for a random sample of 10 from a large consignment gave the following

Item	1	2	3	4	5	6	7	8	9	10
Life in 000' hrs	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accepted the hypothesis that the average lifetime of bulb is 4000 hrs.

Soln:

Null Hypothesis, $H_0: \mu = 4000$ hrs

" There is no significant difference in sample mean and population Mean"

Alternative Hypothesis, $H_1, \mu \neq 4000$ hrs

$t = \bar{x} - \mu$

Alternative Hypothesis, H_1, $\mu \neq 4000$ hrs

Test statistics (t test) : $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$

x	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6
$(x-\bar{x})^2$	0.04	0.04	0.25	0.09	0.64	0.36	0.25	0.01	0	1.44

$\Sigma x = 44$, $n = 10$, , $\bar{x} = \frac{\Sigma x}{n} = 4.4$, $\Sigma(x-\bar{x})^2 = 3.12$

$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 0.589$

thus, $t = \frac{4.4 - 4}{\frac{0.589}{\sqrt{10}}} = 2.123$

t - test table : $\alpha = 0.05$, and $df = 10 - 1 = 9$

thus, $t_{0.05, 9} = 2.26$

Conclusion: $|t| < t_\alpha$

Therefore, Null Hypothesis H_0 is accepted

$\Rightarrow$ there is no significance difference b/w sample mean and hypothetical population Mean.

$\Rightarrow$ Sample have been taken from the Correct population

Table 2 : SIGNIFICANT VALUES $t_v(\alpha)$ OF t-DISTRIBUTION (TWO TAIL AREAS) $[\,|t| > t_v(\alpha)] = \alpha$

d.f. (v)	Probability (Level of Significance) 0.50	0.10	0.05	0.02	0.01	0.001
1	1.00	6.31	12.71	31.82	63.66	636.62
2	0.82	0.92	4.30	6.97	6.93	31.60
3	0.77	2.32	3.18	4.54	5.84	12.94
4	0.74	2.13	2.78	3.75	4.60	8.61
5	0.73	2.02	2.57	3.37	4.03	6.86
6	0.72	1.94	2.45	3.14	3.71	5.96
7	0.71	1.90	2.37	3.00	3.50	5.41
8	0.71	1.80	2.31	2.90	3.36	5.04
9	0.70	1.83	2.26	2.82	3.25	4.78
10	0.70	1.81	2.23	2.76	3.17	4.59
11	0.70	1.80	2.20	2.72	3.11	4.44
12	0.70	1.78	2.18	2.68	3.06	4.32
13	0.69	1.77	2.16	2.05	3.01	4.22
14	0.69	1.76	2.15	2.62	2.98	4.14
15	0.69	1.75	2.13	2.60	2.95	4.07
16	0.69	1.75	2.12	2.58	2.92	4.02
17	0.69	1.74	2.11	2.57	2.90	3.97
18	0.69	1.73	2.10	2.55	2.88	3.92
19	0.69	1.73	2.09	2.54	2.86	3.88
20	0.69	1.73	2.09	2.53	2.85	3.85
21	0.69	1.72	2.08	2.52	2.83	3.83
22	0.69	1.72	2.07	2.51	2.42	3.79
23	0.69	1.71	2.07	2.50	2.81	3.77
24	0.69	1.71	2.06	2.49	2.80	3.75
25	0.68	1.71	2.06	2.49	2.79	3.73
26	0.68	1.71	2.06	2.48	2.78	3.71
27	0.68	1.70	2.05	2.47	2.77	3.69
28	0.68	1.70	2.05	2.47	2.76	3.67
29	0.68	1.70	2.05	2.46	2.76	3.66
30	0.68	1.70	2.04	2.46	2.75	3.65
∞	0.67	1.65	1.96	2.33	2.58	3.29

Two sample t-test :-

(i) To test the difference b/w Mean of two independent sample

(ii) Two sample t- test applicable when

ⓐ Sample sizes $n_1 \leq 30$ & $n_2 \leq 30$

ⓑ The population variance are equal i.e; $\sigma_1^2 = \sigma_2^2$

Working Rule :-

(i) Null Hypothesis, $H_0 : \mu_1 = \mu_2$

"The mean of both population are same"

Alternative Hypothesis, $H_1 \quad \mu_1 \neq \mu_2$

(i) t- test

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

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

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

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

If S.D s_1 & s_2 given

find $|t|$

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

If S.D. s_1 & s_1 are not given

(iii) Using t-test table to find $t_{\alpha, df}$

α = level of significance

df = degree of freedom.

$(n_1 + n_2 - 2)$

(iv) Conclusion

ⓐ if $|t| < t_\alpha$ the H_0 is accepted

ⓑ if $|t| > t_\alpha$ then H_0 is rejected.

Note:- By this test we testing

wither (i) two sample mean $\bar{x}$ & $\bar{y}$ do not differ significantly

or (ii) The two independent sample have been drawn from the population with the same means.

Q. Sample of sizes 10 and 14 were taken from two normal population with from two normal population with S.D 3.5 & 5.2 the sample means were found to be 20.3 and 18.6 test whether the means of the two population are the same at 5% level.

Q.2. The hight of 6 Randomly chosen sailors in inches are 63, 65, 68, 69, 71 and 72. Those of 9 Randomly chosen soldiers are 61, 62, 65, 66, 69, 70, 71, 72 and 73. Test whether the sailors are on the

69, 71 and 72. Those of 9 Randomly chosen soldiers are 61, 62, 65, 66, 69, 70, 71, 72 and 73. Test whether the sailors are on the average taller than soldiers.

Sol^{n}

Null Hypothesis, $H_0: \mu_1 = \mu_2$

"the mean of both the population are the same"

Alternative Hypothesis, $H_1: \mu_1 > \mu_2$ (one tailed test)

Calculation of two sample means:

$$\bar{x} = 68 \quad \& \quad \bar{y} = 67.666$$

{ calculation do self }

$$s = 4.038 \quad , \quad n_1 = 6, \quad n_2 = 9$$

Test statistic:

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{68 - 67.66}{4.038\sqrt{\frac{1}{6} + \frac{1}{9}}}$$

$$= 0.1569$$

Thus, $|t| = 0.1569$

Calculate $t_{\alpha, df}$

$$\alpha = 0.05, \qquad df = n_1 + n_2 - 2 = 15 - 2 = 13$$

$$t_{0.05, 13} = 1.77 \quad \text{(one tailed test)}$$

Conclusion:- Since $|t| < t_{\alpha, df}$

Thus, the null hypothesis H_0 is accepted

Therefore, there is no significant difference b/w their mean.

Hence, "the sailors are not on the average taller than the soldiers."

F-test or Snedecors F- Distribution

or Varience Ratio test.

or Fishers F-test.

F-test Working Rule :-

① Null Hypothesis, $H_0 : \sigma_1^2 = \sigma_2^2$

i.e there is no significant difference b/w population variance"

Alternative Hypothesis, $H_1 : \sigma_1^2 \neq \sigma_2^2$

(ii) F-test (Test statistic)

$$\boxed{\begin{array}{ll} F = \dfrac{S_1^2}{S_2^2} & \text{if } S_1^2 > S_2^2 \\ \text{or } F = \dfrac{S_2^2}{S_1^2} & \text{if } S_2^2 > S_1^2 \end{array}}$$

where

S_1^2 = Sample variance of the sample (x) (first), (n_1)

S_2^2 = sample variance of the sample (y) (second), (n_2)

Such that

if sample S.D not given

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1}, \quad S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1}$$

if sample are given in data form

if sample S.D (s_1 & s_2) are given then

$$S_1^2 = \frac{s_1^2 n_1}{n_1 - 1}, \quad S_2^2 = \frac{s_2^2 n_2}{n_2 - 1}$$

Thus, find F

(iii) From the F-test table calculate

$F_{0.05}$ for $(n_1 - 1)$ & $(n_2 - 1)$ degree of freedom
(at 0.05 level of significance)

(iv) Conclusion :

(iv) Conclusion :

(a) If $f < f_{0.05}$ then H_0 is accepted

(b) If $f > f_{0.05}$ then H_1 is accepted.

Note :- If H_0 accepted then the population variance do not differ siginificantly

$\Rightarrow$ Sample have been taken from the normal population with same variance.

Questions

type I

To test the sample have been drawn from normal population with same variance, f-test only.

type II

To test the samples are drawn from the same normal population

(i) f-test (for equal variance)

(ii) t-test (for equal mean)

Example- Two random samples drawn from 2 Normal populations are as follows;

A	17	27	18	25	27	29	13	17
B	16	16	20	27	26	25	21	

Test whether the samples are drawn from the same normal population.

Sol^{n} :- Part I. f-test :

Null Hypothesis, $H_0 : \sigma_1^2 = \sigma_2^2$

The population variance do not differ significantly

Alternative Hypothesis, $H_1 : \sigma_1^2 \neq \sigma_2^2$

Test statistic :- $F = \frac{S_1^2}{S_2^2}$ if $S_1^2 > S_2^2$

Test statistic :- $F = \frac{s_1^2}{s_2^2}$ if $s_1^2 > s_2^2$ —— (1)

Computation value of s_1^2 and s_2^2

X	Y	$(x-\bar{x})^2$	$(y-\bar{y})^2$
17	16	21.39	
27	16	28.89	
18	20	13.14	
25	27	11.39	
27	26	28.89	
29	25	54.39	
13	21	74.39	
17		21.39	
		=	

$\bar{x} = 21.625$, $\bar{y} = 21.57$

$s_1^2 = \frac{253.87}{7} = 36.267$, $s_2^2 = \frac{\sum(y-\bar{y})^2}{6}$
$= 20.946$

from (1).

$F = \frac{36.267}{30.47} = 1.190$

Now, $F_{0.05\,(7,6)} = 4.21$

Conclusion :-
since $F < F_\alpha$,

Thus, H_0 is accepted

Hence, the variance of both populations significantly equal.

Part II t-test:-

Null Hypothesis, $H_0: \mu_1 = \mu_2$
(The population means are equal)

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

Test of statistic

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

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

$s^2 =$

$s = 5.7067$

$S = 5.7962$

$$= \frac{21.625 - 21.57}{5.7962\sqrt{\frac{1}{7}+\frac{1}{6}}}$$

$$= 0.0167$$

find $t_{\alpha, df}$, $\alpha = 0.05$, $df = 13$

thus, $t_{\alpha, df} = 2.16$

Conclusion, Since $|t| < t_{\alpha}$

Hence, Null Hypothesis is accepted

therefore is no Significant difference b/w the population mean and sample mean.

"we conclude that the two sample have drawn from the same normal population.

Q.2 Two independent sample of sizes 7 and 6 had the following values

Sample A	28	30	32	33	31	29	34
Sample B	29	30	30	24	27	28	

Examine whether the samples have been drawn from normal population having the same variance.

Q.3. The two random sample reveal the following data

Sample no	Size	Mean	Variance
I	16	440	40
II	25	460	42

Test whether the sample come from the same normal population

Table 3 : F-Distribution

Values of F for F-Distributions with 0.05 of the Area in the Right Tail

	Degrees of freedom for numerator																		
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*12*	*15*	*20*	*24*	*30*	*40*	*60*	*120*	*∞*
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	9.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	3.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	3.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.17	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.98	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.94	2.40	2.34	2.28	2.24	2.16	2.29	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.64	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

CHI - SQUARE (χ^2) TEST

χ^2 - Test, chi - square test, chi - square test of Goodness of fit

Non - Parametric [χ^2 - test] $\Rightarrow$ If no information about the population (Normal, Poisson, Binomial)

$\Rightarrow$ χ^2 - test is a distribution free testing method.

Type I - 1 χ^2 - test for goodness of fit

Let

observed frequencies :	$O_1, O_2, O_3, \ldots, O_n$
Expected frequencies :	$E_1, E_2, E_3 \ldots E_n$

such that $\sum O_i = \sum E_i = N$ (Total frequence) with

degree of freedom $(df) = n-1$

Working Rule :-

(i) Null Hypothesis, H_0 : There is no significant difference b/w observed and expected.

(ii) χ^2-test

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

(iii) From the χ^2-test table, calculate

$\chi^2_{0.05}$ for the degree of freedom $(n-1)$.

Conclusion :-

(i) If $\chi^2 < \chi^2_{0.05}$, then H_0 accepted

(ii) If $\chi^2 > \chi^2_{0.05}$ then, H_0 rejected

Note :- If H_0 is accepted then there is no significant difference b/w observed frequency and expected frequency.

$\Rightarrow$ Fit of the data is considered to be good.

$\boxed{\begin{array}{ll} \text{Degree of freedom.} & \text{(i) Binomial Distribution } (n-1) \\ & \text{(ii) Poisson distribution } (n-2) \\ & \text{(iii) Normal distribution } (n-3) \end{array}}$

Condition for applying χ^2-test :-

① The total no. of frequencies (N) should be large $(N \geq 50)$

② The expected frequency E_i of any item should not be less than 5.

② The expected frequency E_i of any item should not be less than 5.

③ χ^2 square test totally depend on degree of freedom
$= n-k$
$=$ total no. of observation $-$ No. of constraints

④ Sum of observed frequency = Sum of expected frequency
$\Rightarrow \Sigma O_i = \Sigma E_i$

Note: If $\chi^2 = 0 \Rightarrow \Sigma O_i = \Sigma E_i$
$\Rightarrow$ the observed and expected frequencies agree exactly.

Ex: The following table gives the number of accident that took place in an industry during various days of week. Test if accidents are uniformly distributed over the week.

Day	Mon	Tue	Wed	Th	Fri	Sat
No. of accidents	14	18	12	11	15	14

Q. A die is thrown 276 times and the results of these throws are given below -

No. appeared on die	1	2	3	4	5	6
frequency	40	32	29	59	57	59

Test wheather the die is biased or not. (AKTU-2019)

Sol^n:- Null Hypothesis, H_0: Die is unbiased

Under this H_0, the expected frequencies for each digit is $\frac{276}{6} = 46$

Find χ^2 as:

O_i	E_i	$(O_i - E_i)^2$
40	46	36
32	46	196
29	46	289
59	46	169
57	46	121
59	46	169

$\Sigma (O_i - E_i)^2 = 980$

$\chi^2 = \frac{980}{46} = 21.30$

Tabulated value of χ^2 at 5% level of Significance for $(6-1=5)$

$\chi^2_{0.05,5} = 11.09$

$\chi^2_{005,5} = 11.09$

Conclusion :- $\chi^2 > \chi^2_{005}$, Thus, H_0 is rejected.
Hence die unbaised or die is biased

Table 4 : CHI-SQUARE

Significant Values χ^2 (α) of Chi-Square Distribution Right Tail Areas for Given Probability α,

$P = P_r (\chi^2 > \chi^2(\alpha)) = \alpha$

And v is Degrees of Freedom (d.f.)

Degree of freedom (v)	*Probability (Level of Significance)*						
	0.99	*0.95*	*0.50*	*0.10*	*0.05*	*0.02*	*0.01*
1	.000157	.00393	.455	2.706	3.841	5.214	6.635
2	.0201	.103	1.386	4.605	5.991	7.824	9.210
3	.115	.352	2.366	6.251	7.815	9.837	11.341
4	.297	.711	3.357	7.779	9.488	11.668	13.277
5	.554	1.145	4.351	9.236	11.070	13.388	15.086
6	.872	2.635	5.348	10.645	12.592	15.033	16.812
7	1.239	2.167	6.346	12.017	14.067	16.622	18.475
8	1.646	2.733	7.344	13.362	15.507	18.168	20.090
9	2.088	3.325	8.343	14.684	16.919	19.679	21.669
10	2.558	3.940	9.340	15.987	18.307	21.161	23.209
11	3.053	4.575	10.341	17.275	19.675	22.618	24.725
12	3.571	5.226	11.340	18.549	21.026	24.054	26.217
13	4.107	5.892	12.340	19.812	22.362	25.472	27.688
14	4.660	6.571	13.339	21.064	23.685	26.873	29.141
15	4.229	7.261	14.339	22.307	24.996	28.259	30.578
16	5.812	7.962	15.338	23.542	26.296	29.633	32.000
17	6.408	8.672	15.338	24.769	27.587	30.995	33.409
18	7.015	9.390	17.338	25.989	28.869	32.346	34.805
19	7.633	10.117	18.338	27.204	30.144	33.687	36.191
20	8.260	10.851	19.337	28.412	31.410	35.020	37.566
21	8.897	11.591	20.337	29.615	32.671	36.343	38.932
22	9.542	12.338	21.337	30.813	33.924	37.659	40.289
23	10.196	13.091	22.337	32.007	35.172	38.968	41.638
24	10.856	13.848	23.337	32.196	36.415	40.270	42.980
25	11.524	14.611	24.337	34.382	37.65	41.566	44.314
26	12.198	15.379	25.336	35.363	38.885	41.856	45.642
27	12.879	16.151	26.336	36.741	40.113	41.140	46.963
28	13.565	16.928	27.336	37.916	41.337	45.419	48.278
29	14.256	17.708	28.336	39.087	42.557	46.693	49.588
30	14.933	18.493	29.336	40.256	43.773	47.962	50.892

Q. The theory predicts the proportion of beans in the fours G_1, G_2, G_3, G_4 should be in the ratio 9:3:3:1. In an experiment with 1600 beans the numbers in the four groups were 882, 313, 287, and 118. Does the experimental result support were 882, 313, 287, and 118. Does the experimental result support the theory.

Sol:-

Null Hypothesis, H_0: The experimental result support the theory.

Under H_0, expected frequencies

$E(G_1) = \frac{1600 \times 9}{16} = 900$, $E(G_2) =$

$E(G_3) =$ $E(G_4) =$

calculate the value of χ^2.

Observed freq O_i	882	313	287	118
Expected freq E_i	900	300	300	100
$(O_i - E_i)^2/E_i$	0.36	0.5633	0.5633	3.24

$$\chi^2 = \Sigma \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$= 4.7266$$

Tabulated value of χ^2 at 0.05 & df = 4-1 = 3

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

Conclusion:- Since $\chi^2 < \chi^2_{0.05}$, Thus null Hypothesis H_0 is accepted

Hence, experimental result support the theory.

Q. Fit a Poisson distribution to the following data and test the goodness of fit

x:	0	1	2	3	4
f:	109	65	22	3	1

Sol:- Null Hypothesis, H_0: Poisson distribution is a good fit to the data

Sol: Null Hypothesis, H_0: Poisson distribution is a good fit to the data

$$\text{Mean}, (\lambda) = \frac{\Sigma fx}{\Sigma f} = \frac{122}{200} = 0.61$$

By Poisson distribution

$$N(x) = 200 \times \frac{e^{-0.61} \times (0.61)^x}{x!}$$

Under H_0, expected frequencies are

$N(0) = 108.67 = 109$ $\qquad$ $N(1) =$

$N(2) =$ $\qquad$ $N(3) =$

$N(4) =$

χ^2 -table is as :

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
109	109	0	0
65	66	1	0.01515
22	20	4	0.2
$\left.\begin{matrix}3\\1\end{matrix}\right\}4$	$\left.\begin{matrix}4\\1\end{matrix}\right\}5$	1	0.2

$= 0.41515$

$\chi^2_{cal} = 0.41515$

and χ^2 at 5% level of significance, degree of freedom $= 5-2-1$
$= 2$

Conclusion:- Since, $\chi^2_{cal} < \chi^2_{tab}$, The null Hypothesis accepted.

Hence, the Poisson distribution is a good fit to the given data.

χ^2 - Test as a test of independence

Degree of freedom $= (r-1)(s-1)$

Q. from the following table regrading the colour of eyes of father and son, test if the colour of son's eye is associated with that the the father

Eye colour of father	Eye colour of son: light	Not light	
light	471	51	= 522
Not light	148	230	= 278
	= 619	= 281	

Sol^{n}: Null Hypothesis, H_0: The colour of son's eye is not associated with that of the father; they are independent

Under H_0, Expected frequencies in each cell = $\dfrac{\text{Product of colomn total and row total}}{\text{whole total}}$

Eye colour of father \ Eye colour of son	light	Not light	Total
light	$\frac{619 \times 522}{900} = 359.02$ a_{11}	162.98 a_{12}	
Not light	259.98 a_{21}	118.02 a_{22}	
Total	619	281	900

χ^2 table as:-

O_i	E_i	$(O_i - E_i)^2 / E_i$
471	359.02	34.92
51		
148		
230		
		$\chi^2 = 261$

find χ^2_{tab}, $d.f = (2-1)(2-1) = 1$ & $\alpha = 0.05$

find χ^2_{tab}, $d.f = (2-1)(2-1) = 1$ & $\alpha = 0.05$

$\chi^2_{tab} = 3.841$

Conclusion :- $\chi^2_{cal} > \chi^2_{tab}$, Hence, Null Hypothesis H_0 rejected.

Therefore the colour of son's eye is associated with that of the father.

Q. From the following data, find wether hair colour and sex are associated

Sex \ Colour	Fair	Red	Medium	Dark	Black	Total
Boys	592	849	504	119	36	2100
Girls	544	677	451	97	14	1783
Total	1136	1526	955	216	50	3883

Q. To test the effectiveness of inoculation against cholera the following table was obtained

	Attacked	Not Attacked	Total
Inoculated	30	160	190
Not inoculated	140	460	600
Total	170	620	790

ANOVA (Analysis of Variance)

ANOVA is statistical tool that can be used for comparision among more than two group's. It is used for testing the hypothesis of equality of more than two normal population means.

Analysis of Variance is carried on the basis of Ratio

The variance Ratio is obtained by dividing the variance b/w sample by variance within sample.

$$\boxed{F = \frac{\text{Variance b/w the sample}}{\text{Variance within the sample}}}$$

This Ratio form is called F- statistic

Technique of ANOVA

- One-way ANOVA — in Syllabus
- Two way ANOVA

One independent factor effect different sample

Working Rule :-

Null Hypothesis, H_0: $\mu_1 = \mu_2 = \mu_3$ i.e., mean of population from which samples are drawn are equal.

Alternative Hypothesis, H_1 : at least one mean is different.

Using table as: (Make) :-

x_i	x_i^2	y_i	y_i^2	z_i	z_i^2
⋮					
Σx_i	Σx_i^2	Σy_i	Σy_i^2	Σz_i	Σz_i^2

* Grand total (GT) = $\Sigma x_i + \Sigma y_i + \Sigma z_i$

* Correction factor $(C.f) = \frac{(G.T.)^2}{\quad}$ where n → Total no.

* Correction factor $(C \cdot f) = \frac{(G \cdot T)^2}{n}$ where $n \rightarrow$ Total no. of items

* Sum of square b/w Sample (SSC)

$$SSC = \frac{(\Sigma x_i)^2}{n_1} + \frac{(\Sigma y_i)^2}{n_2} + \frac{(\Sigma z_i)^2}{n_3} - C \cdot f$$

* Total sum of square (SST)

$$= \Sigma x_i^2 + \Sigma y_i^2 + \Sigma z_i^2$$

* Sum of square within the sample (SSE)

$$\Rightarrow SSE = SST - SSC$$

* Mean sum of square between sample (Variance)

$$MSC = \frac{SSC}{C-1} = \frac{SSC}{df}$$ where C = no. of columns, $df = C-1$

* Mean sum of square within samples (Variance)

$$MSE = \frac{SSE}{n-c} = \frac{SSE}{c(r-1)}$$ where c = no. of column, r = no of Rows.

F - Statistic $\boxed{F_{cal} = \frac{MSC}{MSE}}$

Source of variation	Sum of square	Degree of freedom (df)	Mean square	F
B/w samples (column Means)	SSC	$c-1$	$MSC = \frac{SSC}{c-1}$	$F = \frac{MSC}{MSE}$
Within samples	SSE	$c(r-1)$ or $n-c$	$MSE = \frac{SSE}{n-c}$	
Total	SST	$cr-1$		

Find F_{tab} at 5% level of Significance, $df = c(r-1)$. or $df = n-c$

Conclusion:- $F_{cal} < F_{tab}$ then H_0 accepted

$F_{cal} > F_{tab}$ then H_1 accepted

Example-1 It is desired to compare three hospitals with regards to the no. of deaths per month. A sample of death records were selected from the records of each hospital and the no. of deaths was as given below. From these data suggest a difference in the no. of the deaths per month among three hospitals.

HOSPITALS

A	B	C
3	6	7
4	3	3
3	3	4
5	4	6
0	4	5

Solⁿ:- Null Hypothesis, H_0: there is no diffrence in the no. of deaths per months among three hospitals

Alternate Hypothesis, H_1: There is significant diffrence in the no. of deaths per month among three hospitals.

Level of Significance: we use 5% level of significance

Test Statistic-

A	A^2	B	B^2	C	C^2
3	9	6	36	7	49
4	16	3	9	3	9
3	9	3	9	4	16
5	25	4	16	6	36
0	0	4	16	5	25
=15	=59	=20	=86	=25	=135

Grand total $(G.T) = \Sigma A + \Sigma B + \Sigma C = 15+20+25$
$= 60$

Correction Factor $(C.F) = \dfrac{(G.T)^2}{n}$

$= \dfrac{(60)^2}{}$ where $n = 15$

$$= \frac{(60)^2}{15} \quad \text{where } n = 15$$

$$= 240$$

Total sum of square (SST)

$$= \sum A^2 + \sum B^2 + \sum C^2 - C.F$$

$$= 59 + 86 + 135 - 240$$

$$= 40$$

Sum of square between sample (SSC)

$$= \frac{(\sum A)^2}{n_1} + \frac{(\sum B)^2}{n_2} + \frac{(\sum C)^2}{n_3} - (CF)$$

$$= \frac{(15)^2}{5} + \frac{(20)^2}{5} + \frac{(25)^2}{5} - 240$$

$$= 10$$

Sum of square within samples (SSE) -

$$= SST - SSC$$

$$= 40 - 10 = 30$$

Here, $c = 3$ & $r = 5$

Source of variation	Sum of square	Degree of freedom (df)	Mean square	F
B/w samples	10	2	5	$F = \frac{5}{2.5} = 2$
Within samples	30	12	2.5	
Total	40	14		

Thus, $F_{cal} = 2$. Now find f_{tab}. at 5% df = 2, 12

$f_{tab} = 3.89$

Conclusion - Since $f_{cal} < f_{tab}$

Therefore, the null hypothesis H_0 is accepted

Hence there is no difference in the no. of deaths per month among three hospitals among three hospitals

Q A manufacturing company purchased three new Machines of diffrent makes and wishes to determine wheather one of them is faster than the others in producing a certain output. five hourly production figures are observed at random from each machine and results are given below:

Observations	A_1	A_2	A_3
1	25	31	24
2	30	39	30
3	36	38	28
4	38	42	25
5	31	35	28

Use ANOVA and – 'ne wheather the machines are significantly diffrent in their mean speed

Solⁿ:- Null Hypothesis, H_0: Machine are not Significantly different in thier mean speed i.e; $\mu_1 = \mu_2 = \mu_3$

Alternate Hypothesis, H_1: Machines are Significantly different in their mean speed.

Level of Significance: we use 5% level of significance

Test Statistic-

obervation	A_1	A_1^2	A_2	A_2^2	A_3	A_3^2
1	25	625	31	961	24	576
2	30	900	39		30	900
3	36	1296	38		28	784
4	38	1444	42		25	625
5	31	961	35		28	784

Here $\Sigma A_1 = 160$

$\Sigma A_2 = 185, \Sigma A_3 = 135$

$\Sigma A_1^2 = 5226, \Sigma A_2^2 = 6915$

$\Sigma A_3^2 = 3669$

Grand total (G.T) = $\sum A_1 + \sum A_2 + \sum A_3 = 160 + 185 + 135$
$= 480$

Correction Factor (C.F) = $\frac{(G.T)^2}{n}$

$= \frac{(480)^2}{15}$ where $n = 15$

$= 15360$

Total sum of square (SST)

$= \sum A_1^2 + \sum A_2^2 + \sum A_3^2 - C.F$

$= 5226 + 6915 + 3669 - 15360$

$= 450$

Sum of square between sample (SSC)

$= \frac{(\sum A_1)^2}{n_1} + \frac{(\sum A_2)^2}{n_2} + \frac{(\sum A_3)^2}{n_3} - (CF)$

$= \frac{(160)^2}{5} + \frac{(185)^2}{5} + \frac{(135)^2}{5} - 15360$

$= 250$

Sum of square within samples (SSE) -

$= SST - SSC$

$= 450 - 250 = 200$

Here, $c = 3$ & $r = 5$

Source of variation	Sum of square	Degree of freedom (df)	Mean square	F
B/w samples	250	2	125	$F = \frac{125}{16.67} = 7.49$
Within samples	200	12	16.67	
Total	450	14		

Thus, $F_{cal} = 7.49$ Now find f_{tab}, at 5% $df = 2, 12$

$f_{tab} = 3.89$

Conclusion - Since $f_{cal} < f_{tab}$

Conclusion - Since $f_{cal} < f_{tab}$

Therefore, the null hypothesis H_0 is rejected

Hence the machines are significantly different in thier mean speed.