5. Accepted 6. Rejected 7. Rejected. 8. Rejected.

8.21. SNEDECOR'S VARIANCE RATIO TEST OR F-TEST

Def

(P.T.U., Dec. 2011)

F-Test for equality of population variances

In testing the significance of the difference of two means of two samples, we assumed that the two samples came from the same population or population with equal variance. The object of the F-test is to discover whether two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having the same variance. Hence before applying the t-test for the significance of the difference of two means, we have to test for the equality of population variance by using F-test.

Let x_i ($i = 1, 2, 3, ..., n_1$) and y_j ($j = 1, 2, 3, ..., n_2$) be two independent samples of sizes n_1 and n_2 and variances s_1^2 and s_2^2 drawn from two normal populations with variances σ_1^2 and σ_2^2. We set up the null hypothesis H_0: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ *i.e.*, population variances are similar. Or in other words H_0 is that the two independent estimates of the common population variance do not differ significantly.

Under H_0; the test statistic is

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1) \text{ if } S_1^2 > S_2^2$$

or
$$\frac{S_2^2}{S_1^2} \sim F(n_2 - 1, n_1 - 1) \text{ if } S_2^2 > S_1^2$$

where S_1^2, S_2^2 are unbiased estimates of the common population variance σ^2 and are given by

$$S_1^2 = \frac{\sum_{j=1}^{n_2} (x_i - \bar{x})^2}{n_2 - 1}, S_2^2 = \frac{\sum_{j=1}^{n_2} (y_j - \bar{y})^2}{n_2 - 1}$$

$\bar{x}$ and $\bar{y}$ are means of the two samples.

Since F-test is based on the ratio of two variances, it is also known as Variance Ratio Test.

Note. In numerical problems, usually sample variance s^2 is given from which S^2 can be obtained on using the relation.

$$ns^2 = (n - 1) S^2 \Rightarrow S^2 = \frac{n}{n-1} s^2.$$

Conclusion. If the calculated value of F exceeds $F_{0.05}$ for $((n_1 - 1), (n_2 - 1))$ degrees of freedom given in table we conclude that the ratio is significant at 5% level.

i.e., we conclude that the sample could have come from two normal population with same variance.

The assumptions on which F-test is based are

1. *The populations for each sample must be normally distributed.*
2. *The samples must be random and independent.*
3. *The ratio of σ_1^2 to σ_2^2 should be equal to 1 or greater than 1. That is why we take the larger variance in the Numenator of the ratio.*

Applications. *F-test is used to test (i) whether two independent samples have been drawn from the normal populations with the same variance σ^2.*

(ii) Whether the two independent estimates of the population variance are homogeneous or not.

ILLUSTRATIVE EXAMPLES

Example 1. *In two independent samples of sizes 8 to 10 the sum of squares of deviation of the sample values from the respective sample means were 84.4 and 102.6. Test whether the difference of variances of the populations is significant or not.* **(P.T.U., June 2003, Jan. 2010, May 2010)**

Solution. Given $n_1 = 8, n_2 = 18; \Sigma(x-\bar{x})^2 = 84.4, \Sigma(y-\bar{y})^2 = 102.6$

Null hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ *i.e.*, there is no significant difference between population variance.

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

Now, $$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{84.4}{7} = 12.057; s_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{102.6}{9} = 11.4$$

Under H_0, the test statistic $F = \dfrac{s_1^2}{s_2^2} = \dfrac{12.057}{11.4} = 1.0576\left(s_1^2 > s_2^2\right)$.

Conclusion. The tabulated value of F at 5% level of significance for (7, 9) d.f. is 3.29

$\therefore$ $F_{0.05} = 3.29$ and $|F| = 1.0576 < 3.29 = F_{0.05} \Rightarrow H_0$ is accepted.

$\therefore$ There is no significant difference between the variance of the populations.

Example 2. *Two independent samples of sizes 9 and 8 give the sum of squares of deviations from their respective means equal to 160 inch² and 91 inch² respectively. Can these be regarded as drawn from the same normal population?* **(P.T.U., May 2006)**

Solution. Here $n_1 = 9; n_2 = 8; \Sigma(x-\bar{x})^2 = 160; \Sigma(y-\bar{y})^2 = 91$

$\therefore$ $v_1 =$ d.f. of 1st sample $= 9 - 1 = 8$

$v_2 =$ d.f. of 2nd sample $= 8 - 1 = 7$

H_0 ; $\sigma_1^2 = \sigma_2^2 = \sigma^2$ *i.e.*, samples can be drawn from the same normal population

H_1; $\sigma_1^2 \neq \sigma_2^2$

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{160}{8} = 20$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{91}{7} = 13$$

$\therefore$ Test statistic $$F = \frac{S_1^2}{S_2^2} \quad \because \quad S_1^2 > S_2^2$$

$$= \frac{20}{13} = 1.5385.$$

The tabulated value of F at 5% level of significance *i.e.*, $F_{0.05}$ for $(8,7) = 3.73$.

Conclusion. As calculated value of $|F| = 1.5385 < 3.73$, the tabulated value

$\therefore$ H_0 is accepted.

Hence, samples can not be drawn from the same normal population.

Example 3. *The I.Q's of 25 students from one college showed a variance of 16 and those of an equal number from the other college had a variance of 8. Discuss whether there is any significant difference in variability of intelligence.* **(P.T.U., 2004)**

Solution. Null hypothesis $H_0 ; \sigma_1^2 = \sigma_2^2$ *i.e.*, There is no significant difference in variability of intelligence

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

Given $$n_1 = 25, s_1^2 = 16, n_2 = 25, s_2^2 = 8$$

Here, variances of samples are given ∴ first of all we will find S_1^2 and S_2^2 ∴ $S_1^2 = \frac{n_1}{n_1 - 1} s_1^2 = \frac{25}{24} \cdot 16 = \frac{50}{3}$;

$$S_2^2 = \frac{n_2}{n_2 - 1} s_2^2 = \frac{25}{24} \cdot 8 = \frac{25}{3}$$

As $S_1^2 > S_2^2$ ∴ test statistic $F = \frac{S_1^2}{S_2^2} = \frac{50}{3} \cdot \frac{3}{25} = 2$.

Tabulated value of F [degree of freedom] for $\nu(n_1 - 1, n_2 - 1) = \nu(24, 24)$ at 5% level = 1.98.

Conclusion. Calculated value of F = 2 is slightly > the tabulated value of $F_{0 \cdot 05} = 1 \cdot 98$

Hence, variability of intelligence is just significant at 5% level of significance.

Example 4. *It is desired to find whether there is less variability in the silver plating done by company A than in that done by company B. If the independent random samples of size 12 of two companies work yield $s_1 = 0.035$ and $s_2 = 0.062$. Test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $(\sigma_1^2 < \sigma_2^2)$ at 0.05 level of significance.* **(P.T.U., May 2005)**

Solution. Null hypothesis; $H_0 : \sigma_1^2 = \sigma_2^2$ *i.e.*, there is no significance difference in variabilities of the silver plating of two companies.

$$H_1 : \sigma_1 \neq \sigma_2$$

given $\quad n_1 = 12, s_1 = 0.035$

$\quad n_2 = 12, s_2 = 0.062$

$$S_1^2 = \frac{n_1}{n_1 - 1} s_1^2 = \frac{12}{11}(0.035)^2 ; S_2^2 = \frac{n_2}{n_2 - 1} s_2^2 = \frac{12}{11}(0.062)^2$$

As $\quad S_2^2 > S_1^2$ ∴ test statistic is $F = \frac{S_2^2}{S_1^2} = \frac{\frac{12}{11}(0.062)^2}{\frac{12}{11}(0.035)^2} = \frac{0.00384}{0.00122} = 3.1379$

Tabulated value of F for degree of freedom $\nu(n_2 - 1, n_1 - 1) = \nu(11, 11)$ at 5% level = 2.82.

Conclusion. Calculated value of F = 3.1379 > the tabulated value of $F_{0.05} = 2.82$.

∴ H_0 is rejected.

Hence, there is variability in the silver plating.

Example 5. *Two random samples have the following values:*

Sample 1	15	22	28	26	18	17	29	21	24
Sample 2	8	12	9	16	15	10			

Test the difference of the estimates of the population variances at 5% level of significance. **(P.T.U., May 2008)**

Solution. H_0 : The population variance do not differ significantly, *i.e.*, $\sigma_1^2 = \sigma_2^2$

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

Given $\quad n_1 = 9, n_2 = 6; \bar{x} = \frac{\Sigma x_i}{n_1} = \frac{15 + 22 + 28 + 26 + 18 + 17 + 29 + 21 + 24}{9} = \frac{200}{9} = 22.22$

and $\qquad \bar{y} = \frac{\Sigma y_i}{n_2} = \frac{70}{6} = 11.66.$

As the values of $\bar{x}$ and $\bar{y}$ are fractions and also the values of x and y are not small ∴ to compute S_1^2 and S_2^2 we apply step deviation method (see Note 2 of art 8.19).

Let the assumed mean of sample I be A = 18 and the assumed mean of sample II be B = 12

x	$d_1 = x - A$	d_1^2	y	$d_2 = y - B$	d_2^2
15	– 3	9	8	– 4	16
22	4	16	12	0	0
28	10	100	9	– 3	9
26	8	64	16	4	16
18	0	0	15	3	9
17	–1	1	10	– 2	4
29	11	121			
21	3	9			
24	6	36			
	$\Sigma d_1 = 38$	$\Sigma d_1^2 = 356$		$\Sigma d_2 = -2$	$\Sigma d_2^2 = 54$

Now, $$S_1^2 = \frac{1}{n_1 - 1}\left[\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1}\right] = \frac{1}{8}\left[356 - \frac{(38)^2}{9}\right] = 24.45$$

$$S_2^2 = \frac{1}{n_2 - 1}\left[\Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right] = \frac{1}{5}\left[54 - \frac{4}{6}\right] = 10.67$$

As $\qquad S_1^2 > S_2^2$

∴ Under null hypothesis test statistic is $F = \frac{S_1^2}{S_2^2} = \frac{24.45}{10.67} = 2.291.$

Conclusion. Tabulated value of F at $v_1 = 9 - 1 = 8$ and $v_2 = 6 - 1 = 5$ is 4.82.

Since, calculated value of F is less than the tabulated value. H_0 is accepted.

∴ There is no significance difference between the population variance.

Example 6. *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 populations having the same variance.

(P.T.U., May 2009)

Solution. H_0 : The variances are equal *i.e.*, $\sigma_1{}^2 = \sigma_2{}^2$ *i.e.*, The samples have been drawn from normal populations with same variance.

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

Given $\qquad n_1 = 7, n_2 = 6$

$$\left.\begin{aligned}\bar{x} &= \text{mean of sample A} = \frac{\Sigma x}{n_1} = 31 \\ \bar{y} &= \text{mean of sample B} = \frac{\Sigma y}{n_2} = 28\end{aligned}\right] \text{both } \bar{x}, \bar{y} \text{ are integers}$$

Computations for S_1^2 and S_2^2:

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
28	−3	9	29	1	1
30	−1	1	30	2	4
32	1	1	30	2	4
33	2	4	24	−4	16
31	0	0	27	−1	1
29	−2	4	28	0	0
34	39	9			
		28			26

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1 - 1} = \frac{28}{6} = 4.666$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2 - 1} = \frac{26}{5} = 5.2.$$

As $\quad S_2^2 > S_1^2$

$\therefore$ Under null hypothers the test statistics is $F = \frac{S_2^2}{S_1^2} = \frac{5.2}{4.67} = 1.1158.$

Conclusion. The tabulated value of F at $\nu_1 = 6 - 1$ and $\nu_2 = 7 - 1$ d.f. for 5% level of significance is 4.39.

Since the calculated value of F is less than the tabulated value H_0 is accepted *i.e.*, there is no significant difference between the variance. *i.e.*, the samples have been drawn from the normal population with same variance.

Example 7. *The two random samples reveal the following data:*

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

Test whether the samples come from the same normal population.

Solution. A normal population has two parameters namely the mean μ and the variance σ^2. To test whether the two independent samples have been drawn from the same normal population, we have to test

(*i*) the equality of means (*ii*) the equality of variance

Since the *t*-test assumes that the sample variances are equal, we first apply F-test *i.e.*, prove (*ii*) part first and (*i*) part later.

F-test: Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

The population variance do not differ significantly.

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

Given $\quad n_1 = 16, n_2 = 25, s_1^2 = 40, s_2^2 = 42$

$$\therefore \quad S_1^2 = \frac{n_1}{n_1 - 1} s_1^2 = \frac{16}{15} \cdot 40 = \frac{128}{3}; S_2^2 = \frac{n_2}{n_2 - 1} s_2^2 = \frac{25}{24} \times 42 = \frac{175}{4}.$$

under null hypothesis H_0 ; test statistic is $F = \frac{S_2^2}{S_1^2}\left(S_2^2 > S_1^2\right) = \frac{175}{4} \times \frac{3}{128} = 1.02$.

Conclusion. The calculated value of F is 1.02. The tabulated value of F at (24.15) d.f. for 5% level of significance is 2.23 Since calculated value of F < tabulated value of F $\therefore$ H_0 is accepted $\therefore$ the population variances are equal.

***t*-test: Null hypothesis.** $H_0 : \mu_1 = \mu_2$ *i.e.*, the population means are equal.

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

Given : $n_1 = 16,\ n_2 = 25,\ \bar{x} = 440,\ \bar{y} = 460$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{16 \times 40 + 25 \times 42}{16 + 25 - 2} = 43.333 \quad \therefore \quad S = 6.582$$

under null hypothesis the test statistic, $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{440 - 460}{6 \cdot 582\sqrt{\dfrac{1}{16} + \dfrac{1}{25}}} = -9.490$ for $(n_1 + n_2 - 2)$ d.f.

Conclusion. The calculated value of $|t|$ is $9 \cdot 490$. The tabulated value of t at 39 d.f. for 5% level of significance is 1.96. Since the calculated value is greater than the tabulated value, H_0 is rejected. *i.e.*, there is significant difference between means *i.e.*, $\mu_1 \neq \mu_2$. Since there is significant difference between means, and no significant difference between variance, we conclude that the samples do not come from the same normal population.

Example 8. *Two random samples drawn from two normal populations have the variable values as below:*

Sample I	*19*	*17*	*16*	*28*	*22*	*23*	*19*	*24*	*26*			
Sample II	*28*	*32*	*40*	*37*	*30*	*35*	*40*	*28*	*41*	*45*	*30*	*36*

Obtain the estimate of the variance of the population and test whether the two population have the same variance.

Solution. Let A = 17 be the assumed mean for sample (I) and B = 28 be the assumed mean for sample (II).

Here $n_1 = 9, n_2 = 12$

x	$d_1 = x - 17$	d_1^2	y	$d_2 = y - 28$	d_2^2
19	2	4	28	0	0
17	0	0	32	4	16
16	−1	1	40	12	144
28	11	121	37	9	81
22	5	25	30	2	4
23	6	36	35	7	49
19	2	4	40	12	144
24	7	49	28	0	0
26	9	81	41	13	169
			45	17	289
			30	2	4
			36	8	64
	$\Sigma d_1 = 41$	$\Sigma d_1^2 = 321$		$\Sigma d_2 = 86$	$\Sigma d_2^2 = 964$

Then $$S_1^2 = \frac{1}{n_1 - 1}\left[\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1}\right] = \frac{1}{8}\left[321 - \frac{(41)^2}{9}\right] = 16.778$$

$$S_2^2 = \frac{1}{n_2 - 1}\left[\Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right] = \frac{1}{11}\left[964 - \frac{(86)^2}{12}\right] = 31.606$$ (using Note 2 art. 8.19).

Under Null Hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ *i.e.*, two populations have the same variance.

Test statistic is $$F = \frac{S_2^2}{S_1^2} \qquad (\because\ S_2^2 > S_1^2)$$

$$= \frac{31.606}{16.778} = 1.884.$$

Conclusion. The calculated value of F is 1·884. The tabulated value of F for $(n_2 - 1, n_1 - 1)$ d.f. = (11, 8) d.f. at 5% level of significance is 3·35. Since the calculated value of F is less than the tabulated value, H_0 is accepted. *i.e.*, there is no significant difference between the population variance *i.e.*, the two populations have the same variance.

TEST YOUR KNOWLEDGE

1. From the following two sample values find out whether they have come from the same population:

Sample 1	17	27	18	25	27	29	27	23	17
Sample 2	16	16	20	16	20	17	15	21	

2. The daily wages in Rupees of skilled workers in two cities are as follows:

	Size of sample of workers	*S.D. of wages in the sample*
City A	16	25
City B	13	32

Find out whether the variance of population is significant or not.

3. The standard deviation calculated from two random samples of sizes 9 and 13 are 2·1 and 1·8 respectively. May the samples be regrated as drawn from normal populations with the same standard deviation?

4. Two independent samples of size 8 and 9 had the following values of the variables:

Sample 1	20	30	23	25	21	22	23	24	
Sample 2	30	31	32	34	35	29	28	27	26

Do the estimates of the population variance differ significantly?

5. The mean diameters of rivets produced by two firms A and B are practically the same but the standard deviation may differ. For 22 rivets produced by firm A, the S.D. is 2.9 mm, while for 16 rivets manufactured by firm B, the S.D. is 3.8 mm. Compare the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test its significance. **[Hint:** Consult S.E. 4]

6. A random sample of 16 values from a normal population has a mean of 41.5 inches and sum of the squares of deviations from the mean equal to 135 inches. Another sample of 20 values from an unknown population has a mean of 43.0 inches and sum of the squares of deviations from their mean is equal to 171 inches. Show that the two samples may be regarded as coming from the same normal populations.

[Hint: Consult S.E. 7: $S_1^2 = \frac{135}{15} = 9$; $S_2^2 = 9$

$\therefore$ $F = 1 < 2.23$ H_0 accepted $\therefore \sigma_1 = \sigma_2$, $|t| = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Chi-Square Test

8.22. INTRODUCTION

When a coin is tossed 200 times the theoretical considerations lead us to expect 100 heads and 100 tails, but in practice these results are rarely achieved. To measure the magnitude of discrepancies between theoretical frequencies and observed frequencies, chi-square test is developed. Chi-square is pronounced as 'Ki' square and is represented by the Greek letter χ^2(ki-square).

8.23. DEFINITION OF CHI-SQUARE TEST

Chi-square test measures the degree of discrepancy between observed frequencies and theoretical frequencies *i.e.*, it compares experimentally obtained results with those expected theoretically. It is used as a test statistic in testing a hypothesis that provides a set of theoretical frequencies with which observed frequencies are compared. Thus χ^2 – test is used to determine whether the discrepancy so obtained between observed and theoretical frequencies is due to error of sampling or due to chance.

8.24(a). CHI-SQUARE DISTRIBUTION

The square of a standard normal variate is called a chi-square variate with degree of freedom 1. Thus if x is a random variable following normal distribution with mean μ and standard deviation σ then $\frac{X-\mu}{\sigma}$ is a standard normal variate and $\therefore \left(\frac{X-\mu}{\sigma}\right)^2$ is a chi-square variate with d.f. 1.

If $X_1, X_2, \ldots, X_\nu$ are ν independent random variables following normal distribution with means $\mu_1, \mu_2, \ldots, \mu_\nu$ and standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_\nu$ respectively, then the variate

$$\chi^2 = \left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2 + \ldots + \left(\frac{X_\nu-\mu_\nu}{\sigma_\nu}\right)^2$$

$$= \sum_{i=1}^{\nu}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2$$ which is the sum of the squares of ν independent standard normal variables, follows chi-square distribution with ν d.f.

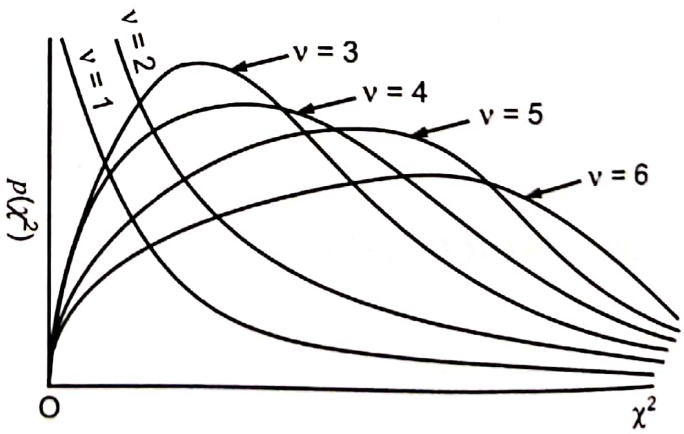
8.24(b). PROBABILITY DENSITY FUNCTION OF CHI-SQUARE DISTRIBUTION

If χ^2 is a random variable following chi-square distribution with ν d.f then its probability function is given by

$$p(\chi^2) = \frac{1}{2^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{\nu}{2}-1};\ 0<\chi^2<\infty$$

where, $\Gamma\left(\frac{\nu}{2}\right)$ is a Gamma function.

Obviously the probability function $p(\chi^2)$ depends upon degree of freedom ν. As ν changes $p(\chi^2)$ changes. The adjoining figure gives the shape of probability curve of chi-square distribution for degree of freedom $\nu = 1, 2, 3, 4, 5, 6$. From the figure it is obvious that as χ^2 increases $p(\chi^2)$ decreases very rapidly and tends to zero as $\chi^2 \to \infty$. So χ^2 axis is an asymptote to the probability curve $p(\chi^2)$.



8.24(c). PROPERTIES OF χ^2-DISTRIBUTION

(P.T.U., May 2009)

1. χ^2 curve is always positively skewed.
2. The number of degrees of freedom is the mean of χ^2-distribution.
3. The value of χ^2 lies between 0 and ∞.
4. The value of χ^2 increases with increasing d.f.
5. The shape of χ^2 curve depends on d.f. but it is not a symmetrical distribution.

8.24(d). APPLICATIONS OF THE χ^2-DISTRIBUTION

Chi-square distribution has a number of applications, some of which are given below:

(*i*) Chi-square test of goodness of fit.

(*ii*) Chi-square test for independence of attributes.

(*iii*) To test if the population has a specified value of the variance σ^2.

(*iv*) To test the equality of several population proportions.

8.25. CHI-SQUARE TEST OF GOODNESS OF FIT

Suppose we are given a set of observed frequencies obtained under some experiment and we want to test if the experimental results support a particular hypothesis or theory. Karl Pearson in 1900, developed a test for testing the significance of the discrepancies between experimental values and the theoretical values obtained under some theory or hypothesis. This test is known as χ^2-test of goodness of fit and is used to test if the deviation between observation (experiments) and theory may be attributed to chance *i.e.*, fluctuation of sampling or if it is really due to the inadequacy of the theory to fit the observed data.

Under null hypothesis H_0; there is no significant difference between observed and the theoretical values.

Let O_i ($i = 1, 2, ..., n$) is a set of observed frequencies and E_i ($i = 1, 2, ..., n$) is the corresponding set of expected frequencies then the test statistic is defined as

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + ... + \frac{(O_n - E_n)^2}{E_n}$$

$$= \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ or simply } \Sigma \frac{(O - E)^2}{E} \text{ with d.f. } \nu = n - 1 \qquad ...(1)$$

Under null hypothesis that the theory fits the data well, look up the tabulated values of χ^2 for $(n - 1)$ d.f at certain level (usually 5% or 1%) from the table given at the end of the book.

(*i*) If the computed value of χ^2 obtained in (1) is < the corresponding tabulated value obtained above, then it is said to be non-significant at the derived level of significance. This implies that the discrepancy between observed and expected frequencies is attributed due to fluctuation of sampling. H_0 is accepted and we conclude that the **fit (correspondence) between theory and experiment is considered to be good.**

(*ii*) On the other hand it calculated value of χ^2 is > the tabulated value then H_0 is rejected and the **fit is considered to be poor** *i.e.*, experiment does not support theory.

Note 1. The observed and the expected frequencies have a very important linear relation:

$$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i = N \text{, The total frequency.}$$

Note 2. If $\chi = 0$, the observed and expected frequencies completely coincide *i.e.*, these agree exactly.

Note 3. In testing goodness of fit the greater the discrepancy between observed and expected frequencies, the greater is the value of χ^2.

Note 4. Since χ^2 does not involve any population parameter, it is termed as statistic and is known as **Non-Parametric Test.**

Note 5. Although χ^2 distribution is a continuous distribution, χ^2-test can be applied to discrete random variables whose frequencies can be counted and tabulated with or without grouping.

For Binomial Distribution $\nu = n - 1$.

For Poisson's Distribution $\nu = n - 2$.

For Normal Distribution $\nu = n - 3$.

ILLUSTRATIVE EXAMPLES

Example 1. *The following table gives the number of accidents that take place in an industry during various days of the week. Test if accidents are uniformly distributed over the week.*

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

Solution. H_0 : Null hypothesis. The accidents are uniformly distributed over the week

Under this H_0, the expected frequencies of the accidents of each of these days $= \frac{84}{6} = 14$

Observed frequency O_i	14	18	12	11	15	14
Expected frequency E_i	14	14	14	14	14	14
$(O_i - E_i)^2$	0	16	4	9	1	0
$\frac{(O_i - E_i)^2}{E_i}$	0	$\frac{16}{14}$	$\frac{4}{14}$	$\frac{9}{14}$	$\frac{1}{14}$	0

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{30}{14} = 2.1428.$$

Conclusion. Table value of χ^2 at 5% level for (6 – 1 = 5 d.f.) is 11.09.

Since, the calculated value of χ^2 is less than the tabulated value H_0 is accepted *i.e.*, the accidents are uniformly distributed over the week.

Example 2. *A die is thrown 276 times and the results of the these throws are given below.*

No appeared on the die	*1*	*2*	*3*	*4*	*5*	*6*
Frequency	*40*	*32*	*29*	*59*	*57*	*59*

Test whether the die is biased or not. **(P.T.U., May 2011)**

Solution. Null hypothesis H_0 : Die is unbiased

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

To find the value of χ^2

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

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{980}{46} = 21.30.$$

Conclusion. Tabulated value of χ^2 at 5% level of significance for (6 – 1 = 5) d.f. is 11.09. Since, the calculated value of $\chi^2 = 21.30 > 11.07$ the tabulated value, H_0 is rejected.

i.e., Die is not unbiased or Die is biased.

Example 3. *The following table shows the distribution of digits in number chosen at random from a telephone directory.*

Digits	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Frequency	*1026*	*1107*	*997*	*966*	*1075*	*933*	*1107*	*972*	*964*	*853*

Test whether the digits may be taken to occur equally frequently in the directory.

Solution. Null hypothesis H_0 : The digits taken in the directory occur equally frequency.

i.e., There is no significance difference between the observed and expected frequency.

Under H_0, the expected frequency is given by $= \frac{10,000}{10} = 1000$

To find the value of χ^2

O_i	1026	1107	997	966	1075	933	1107	972	964	853
E_i	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
$(O_i - E_i)^2$	676	11449	9	1156	5625	4489	11449	784	1296	21609

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{58542}{1000} = 58.542.$$

Conclusion. The tabulated value χ^2 at 5% level of significance for 9 d.f. is 16.919. Since, the calculated value of χ^2 is greater than the tabulated value, H_0 is rejected.

i.e., there is significant difference between the observed and theoretical frequency

i.e., the digits taken in the directory do not occur equally frequently.

Example 4. *Records taken of the number of male and female births in 800 families having four children are as follows :*

No. of male births	*0*	*1*	*2*	*3*	*4*
No. of female births	*4*	*3*	*2*	*1*	*0*
No. of families	*32*	*178*	*290*	*236*	*94*

Test whether the data are consistent with the hypothesis that the binomial law holds and the chance of male birth is equal to that of female birth, namely $p = q = 1/2$. **(P.T.U., Dec. 2006)**

Solution. H_0 : The data are consistent with the hypothesis of equal probability for male and female births. *i.e.* $p = q = 1/2$.

We use binomial distribution to calculate theoretical frequency given by : $N(r) = N \times P(X = r)$ where N is the total frequency. $N(r)$ is the number of families with r male children : $P(X = r) = {}^nC_r p^r q^{n-r}$ where p and q are probability of male and female birth, n is the number of children.

$$N(0) = \text{No. of families with 0 male children} = N \times P(X = 0) = 800 \times {}^4C_0 \left(\frac{1}{2}\right)^4 = 800 \times 1 \times \frac{1}{2^4} = 50$$

$$N(1) = N \times P(X = 1) = 800 \times {}^4C_1 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 200 ; N(2) = N \times P(X = 2) = 800 \times {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = 300$$

$$N(3) = N \times P(X = 3) = 800 \times {}^4C_3 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 = 200 ; N(4) = N \times P(X = 4) = 800 \times {}^4C_4 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 = 50$$

Observed frequency O_i	32	178	290	236	94
Exp-frequency E_i	50	200	300	200	50
$(O_i - E_i)^2$	324	484	100	1296	1936
$\dfrac{(O_i - E_i)^2}{E_i}$	6.48	2.42	0.333	6.48	38.72

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 54.433.$$

Conclusion. Table value of χ^2 at 5% level of significance for $5 - 1 = 4$ d.f. is 9.49.

Since the calculated value χ^2 is greater than the tabulated value, H_0 is rejected.

i.e., the data are not consistent with the hypothesis that the binomial law holds and that the chance of a male birth is not equal to that of a female birth.

Note. Since, the fitting a binomial, the degrees of freedom $v = n - 1$ *i.e.*, $v = 5 - 1 = 4$.

Example 5. *A set of five similar coins is tossed 320 times and the result is*

No. of heads	0	1	2	3	4	5
Frequencies	6	27	72	112	71	32

Test the hypothesis that the data follows a binomial distribution. **(P.T.U., Dec. 2004, May 2009)**

Solution. $p = \text{Probability of heads} = \frac{1}{2}$; $q = \frac{1}{2}$

H_0 ; data follows binomial distribution

H_1 ; data does not follow binomial distribution

We use binomial distribution to calculate theoretical frequency given by

$N(r) = N \times P(X = r)$, where N is the total frequency and $P(X = r) = {}^nC_r q^{n-r} p^r$

Here, $n = 5$

$$N(0) = 320\,P(X=0) = 320 \times {}^5C_0 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = 10$$

$$N(1) = 320\,P(X=1) = 320 \times {}^5C_1 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = 50$$

$$N(2) = 320\,P(X=2) = 320 \times {}^5C_2 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 100$$

$$N(3) = 320\,P(X=3) = 320 \times {}^5C_3 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 100$$

$$N(4) = 320\,P(X=4) = 320 \times {}^5C_4 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 50$$

$$N(5) = 320\,P(X=5) = 320 \times {}^5C_5 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 10$$

O_i	6	27	72	112	71	32
E_i	10	50	100	100	50	10
$(O_i - E_i)^2$	16	529	784	144	441	484
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{16}{10}$	$\frac{529}{50}$	$\frac{784}{100}$	$\frac{144}{100}$	$\frac{441}{50}$	$\frac{484}{10}$

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = \frac{16}{10} + \frac{529}{50} + \frac{784}{100} + \frac{144}{100} + \frac{441}{50} + \frac{484}{10}$$

$$= 78.68$$

Degree of freedom $\nu = 6 - 1 = 5$

Tabulated value of χ^2 at 5% level of significance = 11.07

Since, the calculated value is much greater than the tabulated value, H_0 is rejected.

i.e., data does not follow binomial theorem.

Example 6. *Fit a Poisson distribution to the following data and test for its goodness at 5% level of significance.*

x	*0*	*1*	*2*	*3*	*4*
f	*419*	*352*	*154*	*56*	*19*

(P.T.U., May 2006)

Solution. Given

x	0	1	2	3	4
f	419	352	154	56	19

To fit Poisson distribution we need λ which is mean of the given distribution.

$$\therefore \qquad \lambda = \frac{\Sigma f x}{\Sigma f} = \frac{0 + 1.352 + 2.154 + 3.56 + 4.19}{419 + 352 + 154 + 56 + 19} = \frac{904}{1000} = 0.904$$

Required Poisson distribution is $N \frac{\lambda^r e^{-\lambda}}{r!}$ where $N = \Sigma f = 1000$.

∴ Poisson distribution table is

r	NP(r)	Theoretical frequency
0	$NP(X=0)=1000\frac{e^{-0.904}}{0!}=\frac{1000}{e^{0.904}}=404.95$	405(round off)
1	$NP(X=1)=\frac{1000\,e^{-0.904}(0.904)^1}{1!}=\frac{1000(0.904)}{e^{0.904}}=366.08$	366
2	$NP(X=2)=\frac{1000\,e^{-0.904}(0.904)^2}{2!}=\frac{1000(0.904)^2}{2}=165.468$	165
3	$NP(X=3)=\frac{1000\,e^{-0.904}(0.904)^3}{3!}=\frac{1000(0.904)^3}{6e^{0.904}}=49.86$	50
4	$NP(X=4)=\frac{1000\,e^{-0.904}(0.904)^4}{4!}=\frac{1000(0.904)^4}{24\,e^{0.904}}=11.26$	$11+3=14$ (3 is add to make total 1000)

Test of goodness of fit

Assume H_0 ; Poisson distribution can be fitted to the data

H_1; Poisson distribution can not be fitted to data

X	0	1	2	3	4
O_i	419	352	154	56	19
E_i	405	366	165	50	14
$(O_i-E_i)^2$	196	196	121	36	25
$\frac{(O_i-E_i)^2}{E_i}$	$\frac{196}{405}=0.48$	$\frac{196}{366}=0.53$	$\frac{121}{165}=0.73$	$\frac{36}{50}=0.72$	$\frac{25}{14}=0.127$

$$\chi^2=\Sigma\frac{(O_i-E_i)^2}{E_i}=0.48+0.53+0.73+0.72+0.127$$

$$=2.587$$

For fitting poissons distribution $v=n-2=5-2=3$ d.f. The tabulated value of ψ^2 at 5% level of significance for $v=3$ d.f. is $\psi_{0.05}=7.815$.

Conclusion. The calculated value of χ^2 *i.e.*, $2.587<$ the tabulated value 7.815 at 5% level of significance

∴ H_0 is accepted

∴ Poisson distribution can be best fitted the data.

Example 7. *Verify whether Poisson distribution can be assumed from the data given below :*

No. of defects	*0*	*1*	*2*	*3*	*4*	*5*
Frequency	*6*	*13*	*13*	*8*	*4*	*3*

(P.T.U., May 2007)

Solution. H_0 : Poisson fit is a good fit to the data

Mean of the given distribution $= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{94}{47} = 2$

To fit a Poisson distribution we require λ. Parameter $\lambda = \bar{x} = 2$

We use Poissons distribution to calculate theroetical frequencies given by $N(r) = N \times P(X = r)$, where N is the total frequency and $P(X = r)$ is $\frac{e^{-\lambda}\lambda^r}{r!}$.

By Poisson distribution the frequency of r success is

$$N(r) = N \times \frac{e^{-\lambda}\lambda^r}{r!}, \text{ N is the total frequency}.$$

$$N(0) = 47 \times e^{-2} \cdot \frac{(2)^0}{0!} = 6{\cdot}36 \approx 6; \quad N(1) = 47 \times e^{-2} \cdot \frac{(2)^1}{1!} = 12{\cdot}72 \approx 13$$

$$N(2) = 47 \times e^{-2} \cdot \frac{(2)^2}{2!} = 12{\cdot}72 \approx 13; \quad N(3) = 47 \times e^{-2} \cdot \frac{(2)^3}{3!} = 8{\cdot}48 \approx 9$$

$$N(4) = 47 \times e^{-2} \cdot \frac{(2)^4}{4!} = 4{\cdot}24 \approx 4; \quad N(5) = 47 \times e^{-2} \cdot \frac{(2)^5}{5!} = 1{\cdot}696 \approx 2.$$

X	0	1	2	3	4	5
O_i	6	13	13	8	4	3
E_i	6.36	12.72	12.72	8.48	4.24	1.696
$\frac{(O_i - E_i)^2}{E_i}$	0.2037	0.00616	0.00616	0.02716	0.0135	1.0026

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 1.2864.$$

To fit poisson's distribution d.f $= 6 - 2 = 4$.

Conclusion. The calculated value of χ^2 is 1.2864. Tabulated value of χ^2 at 5% level of significance for $\upsilon = 6 - 2 = 4$ d.f. is 9.49. Since the calculated value of χ^2 is less than that of tabulated value. H_0 is accepted *i.e.*, Poisson distribution provides a good fit to the data.

Example 8. *The theory predicts the proportion of beans in the four groups, 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 the theory.*

Solution. H_0, the experimental result support the theory *i.e.*, there is no significant difference between the observed and theoretical frequency under H_0. The theoretical frequency can be calculated as follows :

$$E(G_1) = \frac{1600 \times 9}{16} = 900; \; E(G_2) = \frac{1600 \times 3}{16} = 300;$$

$$E(G_3) = \frac{1600 \times 3}{16} = 300; \; E(G_4) = \frac{1600 \times 1}{16} = 100$$

To calculate the total value of χ^2.

Observed frequency O_i	882	313	287	118
Exp. frequency E_i	900	300	300	100
$\frac{(O_i - E_i)^2}{E_i}$	0.36	0.5633	0.5633	3.24

$$\chi^2 = \sum \frac{(0_i - E_i)^2}{E_i} = 4.7266.$$

Conclusion. Table value of χ^2 at 5% level of significance for 3 d.f. is 7.815. Since, the calculated value of χ^2 is less than that of the tabulated value. Hence, H_0 is accepted.

i.e., The experimental result support the theory.

Example 9. *The figures given below are (a) the frequencies of a distribution and (b) the frequencies of*