6.8.1 Testing the ratio of variances: (F test for variance ratio)

Suppose we are interested to test whether the two normal population have same variance or not. Let $x_1, x_2, x_3 \ldots\ldots x_{n_1}$, be a random sample of size n_1, from the first population with variance σ_1^2 and $y_1, y_2, y_3 \ldots y_{n_2}$, be random sample of size n_2 form the second population with a variance σ^2. Obviously the two samples are independent.

UNIT-IV 1

Null hypothesis:

$H_0 = \sigma_1^2 = \sigma_2^2 = \sigma^2$

i.e. population variances are same. In other words H_0 is that the two independent estimates of the common population variance do not differ significantly.

Calculation of statistics:

Under H_0, the test statistic is

$$F_0 = \frac{S_1^2}{S_2^2}$$

Where $S_1^2 = \frac{1}{n_1 - 1}\Sigma(x - \bar{x})^2 = \frac{n_1 s_1^2}{n_1 - 1}$

$$S_2^2 = \frac{1}{n_2 - 1}\Sigma(y - \bar{y})^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

It should be noted that numerator is always greater than the denominator in F-ratio

$$F = \frac{\text{Larger Variance}}{\text{Samller Variance}}$$

v_1 = d.f for sample having larger variance

v_2 = d.f for sample having smaller variance

Expected value :

$F_e = \frac{S_1^2}{S_2^2}$ follows F- distribution with $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ d.f

2

The calculated value of F is compared with the table value for v_1 and v_2 at 5% or 1% level of significance If $F_0 > F_e$ then we reject H_0. On the other hand if $F_0 < F_e$ we accept the null hypothesis and it is a inferred that both the samples have come from the population having same variance.

Since F- test is based on the ratio of variances it is also known as the variance Ratio test. The ratio of two variances follows a distribution called the F distribution named after the famous statisticians R.A. Fisher.

Example 14:

Two random samples drawn from two normal populations are :

Sample I:	20	16	26	27	22	23	18	24	19	25		
Sample II:	27	33	42	35	32	34	38	28	41	43	30	37

Obtain the estimates of the variance of the population and test 5% level of significance whether the two populations have the same variance.

Solution:

Null Hypothesis:

$H_0: \sigma_1^2 = \sigma_2^2$ i.e. The two samples are drawn from two populations having the same variance.

Alternative Hypothesis:

$H_1: \sigma_1^2 \neq \sigma_2^2$ (two tailed test)

$$\overline{x}_1 = \frac{\sum x_1}{n_1}$$

$$= \frac{220}{10}$$

$$= 22$$

$$\overline{x}_2 = \frac{\sum x_2}{n_2}$$

$$= \frac{420}{12}$$

$$= 35$$

x_1	$x_1 - \overline{x}_1$	$(x_1 - \overline{x}_1)^2$	x_2	$x_2 - \overline{x}_2$	$(x_2 - \overline{x}_2)^2$
20	−2	4	27	−8	64
16	−6	36	33	−2	4
26	4	16	42	7	49
27	5	25	35	0	0
22	0	0	32	−3	9
23	1	1	34	−1	1
18	−4	16	38	3	9
24	2	4	28	−7	49
19	−3	9	41	6	36
25	3	9	43	8	64
220	**0**	**120**	30	−5	25
			37	2	4
			420	**0**	**314**

Level of significance :

0.05

The statistic F is defined by the ratio

$$F_0 = \frac{S_1^2}{S_2^2}$$

Where $S_1^2 = \frac{\sum(x_1 - \overline{x}_1)^2}{n_1 - 1} = \frac{120}{9} = 13.33$

$$S_2^2 = \frac{\sum(x_2 - \overline{x}_2)^2}{n_2 - 1} = \frac{314}{11} = 28.54$$

Since $S_2^2 > S_1^2$ larger variance must be put in the numerator and smaller in the denominator

$$\therefore F_0 = \frac{28.54}{13.33} = 2.14$$

Expected value:

$F = \frac{S_2^2}{S_1^2}$ follows F- distribution with

$v_1 = 12-1 = 11$; $v_2 = 10-1 = 9$ d.f = 3.10

Inference :

Since $F_0 < Fe$ we accept null hypothesis at 5% level of significance and conclude that the two samples may be regarded as drawn from the populations having same variance.

Example 15:

The following data refer to yield of wheat in quintals on plots of equal area in two agricultural blocks A and B Block A was a controlled block treated in the same way as Block B expect the amount of fertilizers used.

	No of plots	Mean yield	Variance
Block A	8	60	50
Block B	6	51	40

Use F test to determine whether variance of the two blocks differ significantly?

Solution:

We are given that

$n_1 = 8 \quad n_2 = 6 \quad \bar{x}_1 = 60 \quad \bar{x}_2 = 51 \quad s_1^2 = 50 \quad s_2^2 = 40$

Null hypothesis:

$H_0: \sigma_1^2 = \sigma_2^2$ ie there is no difference in the variances of yield of wheat.

Alternative Hypothesis:

$H_1: \sigma_1^2 \neq \sigma_2^2$ (two tailed test)

Level of significance:

Let $\alpha = 0.05$

Calculation of statistic:

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{8 \times 50}{7}$$
$$= 57.14$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{6 \times 40}{5}$$
$$= 48$$

5

Since $S_1^2 > S_2^2$

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{57.14}{48} = 1.19$$

Expected value:

$F_e = \dfrac{S_1^2}{S_2^2}$ follows F- distribution with $v_1 = 8-1 = 7$ $v_2 = 6-1 = 5$ d.f

$= 4.88$

Inference:

Since $F_0 < F_e$, we accept the null hypothesis and hence infer that there is no difference in the variances of yield of wheat.

χ^2-Distribution:- ~~UNIT-IV~~ 6

1. The square of a standard normal variate is know as a χ^2 variate with a one degree of freedom then $X \sim N(\mu, \sigma^2)$

$$\text{when } z = \left(\frac{x-\mu}{\sigma}\right) \sim N(0,1)$$

Therefore $\chi^2 = \left(\frac{x-\mu}{\sigma}\right)^2$ is a χ^2 variate with one degree of freedom.

2. In genral if x_i $(i=1,2,\ldots n)$ are independent normal variates with mean μ_i and variance σ_i^2, $(i=1,2,\ldots n)$ therefore $\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$ is χ^2 variate with n degrees of freedom.

Applications of χ^2-Distribution:- 5

1. To test if the hypothetical value of the population variance is $\sigma^2 = \sigma_0^2$
2. To test the goodness of fit
3. To test the independents of attributes
4. To test homogenity of independent estimates of the population variance
5. To test homogenity of independent estimate of the population correlation co-efficient.

Explain χ^2 test For Goodness fit:-

A very powerful test for testing the significance of the discrepancy describing between theory and Expriment was given Prof karl-pearson is 1960 and is known as χ^2 test of goodness of fit It enable us to find if the deviation of the Experiment from theory is just by chance or is it rearly due to the in adaquacy of the theory.

If $f_i (i = 1, 2, \dots n)$ is a set of observed [Experimental] observe frequence and $e_i (i = 1, 2, \dots n)$ is the corresponding set of expected frequences, then karl-pearson χ^2 is given by $\chi^2 = \sum_{i=1}^{n} \left[\frac{(f_i - e_i)^2}{e_i} \right]$

$\Sigma f_i = \Sigma e_i$ follows χ^2 distribution with $n-1$ degrees of freedom.

—x—

4 7

6.6 Test of independence

Let us suppose that the given population consisting of N items is divided into r mutually disjoint (exclusive) and exhaustive classes $A_1, A_2, \ldots, A_r$ with respect to the attribute A so that randomly selected item belongs to one and only one of the attributes $A_1, A_2, \ldots, A_r$ Similarly let us suppose that the same population is divided into c mutually disjoint and exhaustive classes $B_1, B_2, \ldots, B_c$ w.r.t another attribute B so that an item selected at random possess one and only one of the attributes $B_1, B_2, \ldots, B_c$. The frequency distribution of the items belonging to

8

the classes $A_1, A_2, \ldots, A_r$ and $B_1, B_2, \ldots, B_c$ can be represented in the following $r \times c$ manifold contingency table.

r × c manifold contingency table

B / A	B_1	B_2	…	B_j	…	B_c	Total
A_1	(A_1B_1)	(A_1B_2)	…	(A_1B_j)	…	(A_1B_c)	(A_1)
A_2	(A_2B_1)	(A_2B_2)	…	(A_2B_j)	…	(A_2B_c)	(A_2)
.	.	.	.	.	.	.	.
A_i	(A_iB_1)	(A_iB_2)	…	(A_iB_j)	…	(A_iB_c)	(A_i)
.	.	.	.	.	.	.	.
A_r	(A_rB_1)	(A_rB_2)	…	(A_rB_j)	…	(A_rB_c)	(A_r)
Total	(B_1)	(B_2)	…	(B_j)	…	(B_c)	$\Sigma Ai = \Sigma Bj = N$

(A_i) is the number of persons possessing the attribute A_i, $(i=1,2,\ldots r)$, (Bj) is the number of persons possing the attribute B_j, $(j=1,2,3,\ldots.c)$ and $(A_i\ B_j)$ is the number of persons possessing both the attributes A_i and B_j $(i=1,2,\ldots r\ ,j=1,2,\ldots c)$.

Also $\Sigma A_i = \Sigma B_j = N$

Under the null hypothesis that the two attributes A and B are independent, the expected frequency for (A_iB_j) is given by

$$= \frac{(Ai)(Bj)}{N}$$

Calculation of statistic:

Thus the under null hypothesis of the independence of attributes,the expected frequencies for each of the cell frequencies of the above table can be obtained on using the formula

$$\chi_0^2 = \Sigma\left(\frac{(O_i - E_i)^2}{E_i}\right)$$

6.6.1 2×2 contingency table : 9

Under the null hypothesis of independence of attributes, the value of χ^2 for the 2×2 contingency table

			Total
	a	b	a+b
	c	d	c+d
Total	a+c	b+d	N

is given by

$$\boxed{\chi_0^{\,2} = \frac{N(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}}$$

6.6.2 Yate's correction

In a 2×2 contingency table, the number of d.f. is (2−1)(2−1) =1. If any one of the theoretical cell frequency is less than 5,the use of pooling method will result in d.f = 0 which is meaningless. In this case we apply a correction given by F.Yate (1934) which is usually known as "Yates correction for continuity". This consisting adding 0.5 to cell frequency which is less than 5 and then adjusting for the remaining cell frequencies accordingly. Thus corrected values of χ^2 is given as

$$\chi^2 = \frac{N\left[(a \mp \frac{1}{2})(d \mp \frac{1}{2}) - (b \pm \frac{1}{2})(c \pm \frac{1}{2})\right]^2}{(a + c)(b + d)(a + b)(c + d)}$$

10

Example 9:

1000 students at college level were graded according to their I.Q. and the economic conditions of their homes. Use χ^2 test to find out whether there is any association between economic condition at home and I.Q.

Economic Conditions	IQ		Total
	High	Low	
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

A

B

Solution:

Null Hypothesis:

There is no association between economic condition at home and I.Q. i.e. they are independent.

$$E_{11} = \frac{(A)(B)}{N} = \frac{600 \times 700}{1000} = 420$$

The table of expected frequencies shall be as follows.

			Total
	420	180	600
	280	120	400
Total	700	300	1000

Observed Frequency O	Expected Frequency E	$(O-E)^2$	$\left(\frac{(O-E)^2}{E}\right)$
460	420	1600	3.81
240	280	1600	5.714
140	180	1600	8.889
160	120	1600	13.333
			31.746

$$\chi_o^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 31.746$$

Expected Value:

$\chi_e^2 = \Sigma\left(\frac{(O - E)^2}{E}\right)$ follow χ^2 distribution with (2–1) (2–1) = 1 d.f

= 3.84

Inference :

$\chi_o^2 > \chi_e^2$, hence the hypothesis is rejected at 5 % level of significance. ∴ there is association between economic condition at home and I.Q.

Example 10:

Out of a sample of 120 persons in a village, 76 persons were administered a new drug for preventing influenza and out of them, 24 persons were attacked by influenza. Out of those who were not administered the new drug .12 persons were not affected by influenza.. Prepare

(i) 2x2 table showing actual frequencies.

(ii) Use chi-square test for finding out whether the new drug is effective or not.

Solution:

The above data can be arranged in the following 2 x 2 contingency table.

Table of observed frequencies

New drug	Effect of Influenza		Total
	Attacked	Not attacked	
Administered	24	76 – 24 = 52	76
Not administered	44 –12 = 32 120 -76	12	120 – 76 = 44
Total	120 – 64 = 56 24 + 32 = 56	52 + 12 = 64	120

Null hypothesis:

Attack of influenza' and the administration of the new drug are independent.

12

Computation of statistic:

$$\chi_o^2 = \frac{N(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

$$= \frac{120(24 \times 12 - 52 \times 32)^2}{56 \times 64 \times 76 \times 44}$$

$$= \frac{120(-1376)^2}{56 \times 64 \times 76 \times 44} = \frac{120(1376)^2}{56 \times 64 \times 76 \times 44}$$

$= \text{Anti log } [\log 120 + 2\log 1376 - (\log 56 + \log 64 + \log 76 + \log 44)]$

$= \text{Antilog } (1.2777) = 18.95$

Expected value:

$\chi_e^2 = \Sigma \left(\frac{(O - E)^2}{E} \right)$ follows χ^2 distribution with $(2-1) \times (2-1)$ d.f

$= 3.84$

Inference:

Since $\chi_o^2 > \chi_e^2$, H_0 is rejected at 5 % level of significance. Hence we conclude that the new drug is definitely effective in controlling (preventing) the disease (influenza).