PROBABILITY AND STATISTICS (PMA303)

Lecture-[32]

(F-test for comparison of variances)

Para. & Non-para., Hypothesis Testing: (Unit VI-VII)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Dr. Rajanish Rai
Assistant Professor
Department of Mathematics
Thapar Institute of Engineering and Technology, Patiala

Rajmish Rai

Variance Ratio or F-test for comparison of variances

* In testing the significance of the difference of two means of two means of two samples, we assumed that the two samples come from the same population or population with equal variance.

* F-test is used either for testing the hypothesis about the equality of two population variances or the equality of two or more population means.

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

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

* R.A. Fisher who introduced the term 'variance' in the analysis of statistical data in 1920.

- F-test is based on the ratio of two variances, it is also known as variance ratio test.

Let x_i $(i = 1, 2, 3, \ldots n_1)$ and y_j $(j = 1, 2, 3, \ldots, n_2)$ be two independent random sample (with means $\bar{x}$ and $\bar{y}$, resp.) drawn from normal populations with the same variance.

Let

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2, \quad S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(y_j-\bar{y})^2.$$

The F-statistics is defined by the relation

$$F = \frac{S_1^2}{S_2^2} \quad \text{where } S_1^2 > S_2^2. \quad \text{—— (1)}$$

Numerator should always be more than denominator.

In case $S_2^2 > S_1^2$, then we have

$$\boxed{F = \frac{S_2^2}{S_1^2}.} \quad S_2^2 > S_1^2 \quad \text{—— (2)}$$

In the first case, we say that F has (n_1-1, n_2-1) degree of freedom.

And in the second case, we say that F has (n_2-1, n_1-1) degree of freedom.

(*) The calculating value of F is compared with the table value for (n_1-1, n_2-1) or (n_2-1, n_1-1) as the case may be at 5% or 1% level of significance. If the calculated value of F is greater than the table value then the F ratio is considered significant and the null hypothesis is rejected.

(*) If the calculated value of F is less than the tabulated value, the null hypothesis is accepted and it is inferred that both the samples have come from the population having the same varience.

Assumptions -:

(i) Independent random samples are drawn from each of two normal populations

(ii) The populations for each sample must be normaley distributed.

(iii) The variability of the measurements in the two populations is same and can be measured by a common varience σ^2, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

(iv) The ratio of σ_1^2 to σ_2^2 should be greater than or equal to 1, since larger value from S_1^2 and S_2^2 is taken in the numerator.

Question (1) The time taken by workers in performing a job by method I and method II is given below

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

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

Solution:- From the given data

$n_1 = 6$ and $n_2 = 7$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{134}{6} = 22.3$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{241}{7} = 34.4.$$

x	y	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
20	27	-2.3	-7.4	5.29	54.76
16	33	-6.3	-1.4	39.69	1.96
26	42	3.7	7.6	13.69	57.76
27	35	4.7	0.6	22.09	0.36
23	32	0.7	-2.4	0.49	5.76
22	34	-0.3	-0.4	0.09	0.16
	38		3.6		12.96
$\sum x = 134$	$\sum y = 241$			$\sum(x_i-\bar{x})^2 = 81.34$	$\sum(y_i-\bar{y})^2 = 133.72$

$$\sum (x_i-\bar{x})^2 = 81.34.$$

$$\sum (y_i-\bar{y})^2 = 133.72.$$

$$S_1^2 = \frac{\sum (x_i-\bar{x})^2}{n_1-1} = \frac{81.34}{5} = 16.26$$

and

$$S_2^2 = \frac{\sum (y_i-\bar{y})^2}{n_2-1} = \frac{133.72}{6} = 22.29.$$

and let the Null hypothesis H_0 :

$$\sigma_1^2 = \sigma_2^2$$

i.e., there is no significant difference between the two variances.

Alternative hypothesis H_1 :

$\sigma_1^2 \neq \sigma_2^2$ (there is a significant difference between two variances).

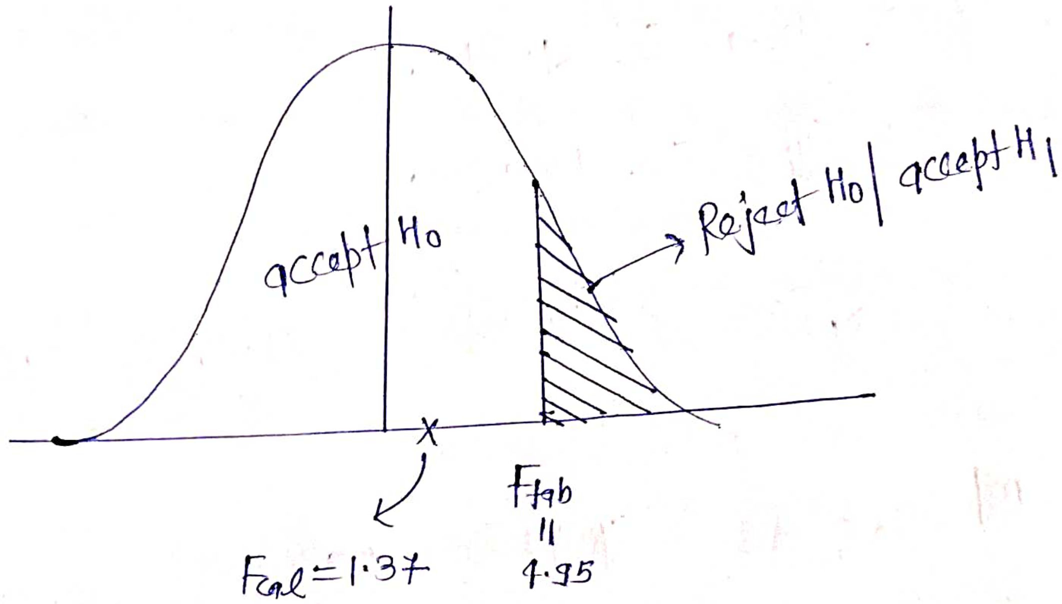
Level of significance $\alpha = 0.05$

$$F(6,5) \text{ at } 0.05 = 4.95$$

Now the test statistics

$$F = \frac{S_2^2}{S_1^2} = \frac{22.29}{16.26} = 1.37$$

$$\boxed{F_{cal} = 1.37}$$



Since $F_{cal} = 1.37 < F_{tab} = 4.95$ at 5% level of significance. So we need not reject the null hypothesis H_0. Hence, we conclude that there is no significant difference between, the two variances at 5% level of significance.

Question:- In a sample of 8 observations, the sum of square of deviations from means is 94.5. In other sample of 10 observations, the sum of square of deviations from mean is 101.7. Test whether there is a significant difference of Variance.

Solution:- Given that

$$n_1 = 8 \quad \text{and} \quad \sum (x_i - \bar{x})^2 = 94.5$$

$$n_2 = 10 \quad \& \quad \sum (y_i - \bar{y})^2 = 101.7.$$

Null hypothesis H_0 : $\sigma_1^2 = \sigma_2^2$ (two workers are equaly stable)

and Alternative Hypothesis

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

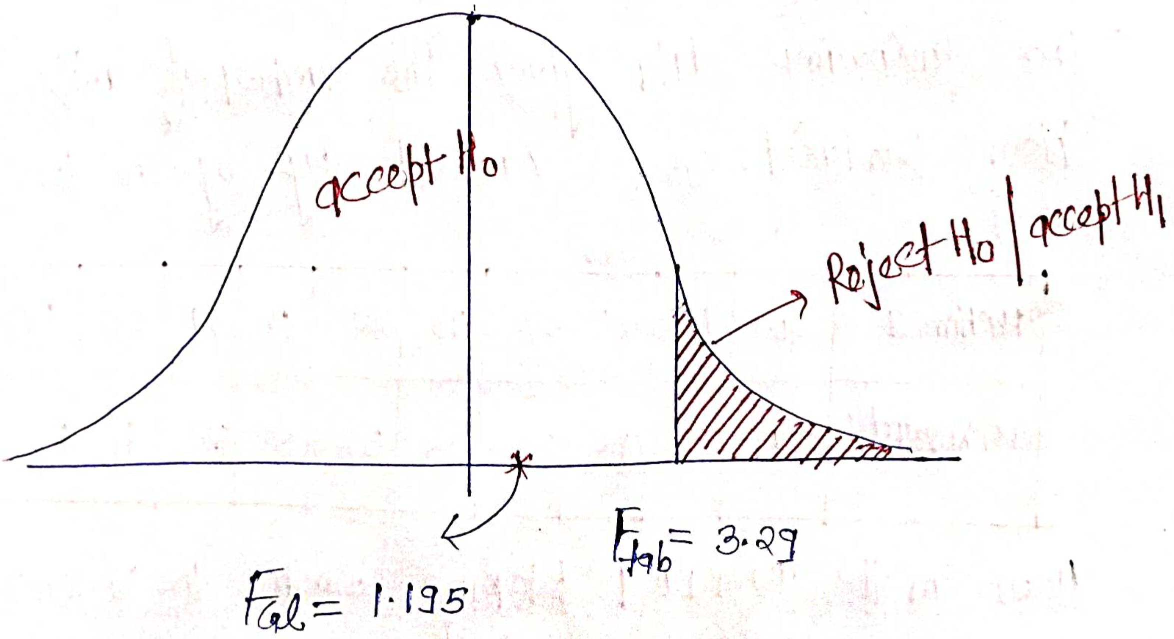
$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{8} (x_i - \bar{x})^2 = \frac{1}{7} \times 94.5 = 13.5$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum (y_i - \bar{y})^2 = \frac{1}{9} \times 101.7 = 11.3$$

$$F = \frac{S_1^2}{S_2^2} = \frac{13.5}{11.3} = 1.195$$

and d.f. $= (8-1, 10-1) = (7, 9)$, the tabulated value

$$F_{0.05}(7, 9) = 3.29$$



As the calculated value of $F_{cal} = 1.94 < F_{0.05}(7,9) = 3.29$, hence hypothesis H_0 is accepted, i.e., the two samples represent the same variance.

Question-: A plant has installed two machines producing polythene bags. During the installation, the manufacturer of the machine has started that the capacity of the machine is to produce 20 bags in a day. Owing to various factor such as different operators working on these machines, raw material, etc., there is a variation in the number of bags produced at the end of the day. The company researcher has taken a random sample of bags produced in 10 days for machines 1 and 13 days for machine 2, respectively. The following data gives the number of units of an item produced on a sampled day by the two machines:

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

How can the researcher determine whether the variance is from same population (population variance are equal) or it comes from different populations (population variances are not equal)? Use 5% level of significance.

Solution :-

(a) Null hypothesis H_0 :- $\sigma_1^2 = \sigma_2^2$

i.e., there is no significant difference between the production capacity of the two machines.

Given that $n_1 = 10$ & $n_2 = 12$

$$\bar{x} = \frac{\sum x}{n} = \frac{220}{10} = 22$$

$$\bar{y} = \frac{\sum y}{n} = \frac{420}{12} = 35.$$

	Machine I		Machine II	
	x	$(x-\bar{x})^2$	y	$(y-\bar{y})^2$
	20	4	27	64
	16	36	33	4
	26	16	42	49
	27	25	35	0
	23	1	32	9
	22	0	34	1
	18	16	38	9
	24	4	28	49
	25	9	41	36
	19	9	43	64
			30	25
			37	4
Total	220	120	420	314

$$S_1^2 = \frac{1}{n_1-1}\sum (x_i-\bar{x})^2 = \frac{120}{10-1} = 13.33$$

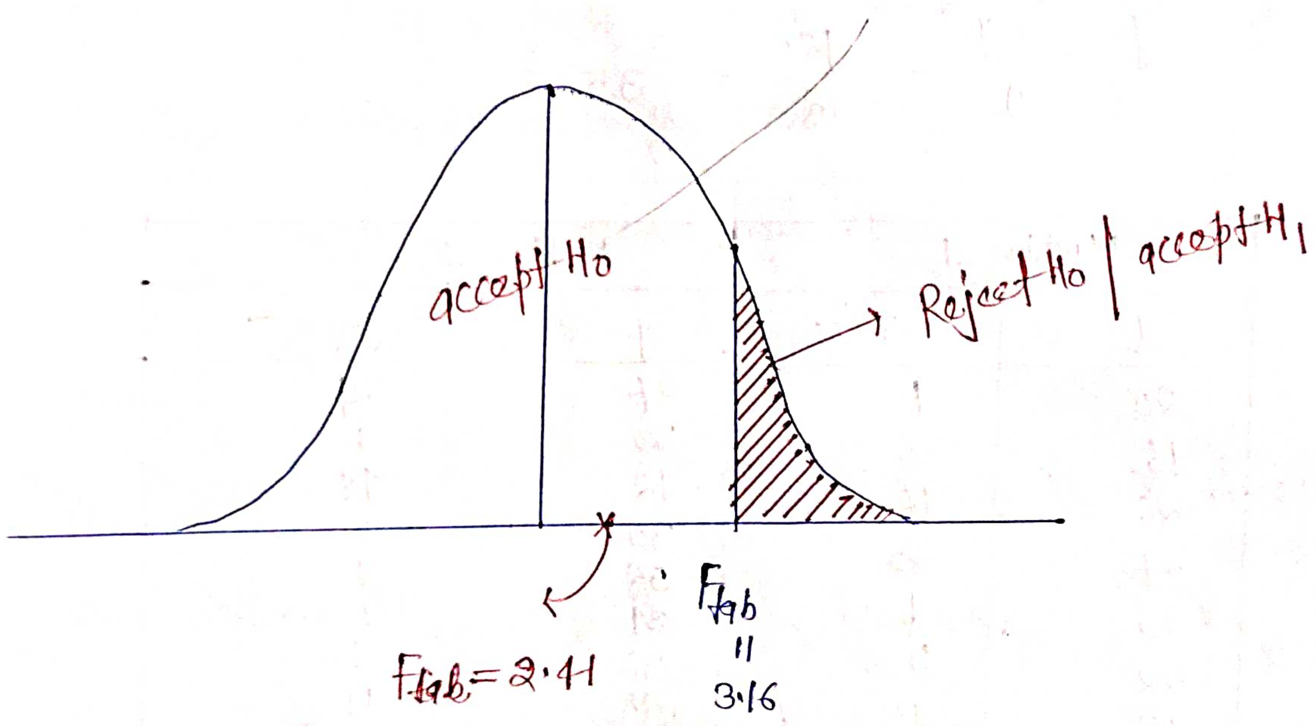
$$S_2^2 = \frac{1}{n_2-1}\sum (y_i-\bar{y})^2 = \frac{314}{12-1} = 28.55$$

$\therefore$ F test $\quad F = \frac{S_2^2}{S_1^2} = \frac{28.55}{13.33}$

$F_{cal} = 2.14$

Step ③ — d.f. $= (12-1, 10-1) = (11, 9)$, the tabulated value at 5% level of significance

$F_{0.05}(11, 9) = 3.16$.



As $F_{cal} = 2.41 < F_{tab} = 3.16$, so Ho is accepted, i.e. there is no significant difference between the production capacity of the two machines.

Question— The daily wages (in ₹) of workers in two cities are as follows:

	Size of the sample	Standard deviation of wages
City (A)	22	2.9
City (B)	16	3.8

Test at 5% level, the equality of variances of the wage distribution in the two cities.

Solution:- Given that

$n_1 = 22$ & $n_2 = 16$

$s_1 = 2.9$ & $s_2 = 3.8$.

(i) Null Hypothesis H_0 :- $\sigma_1^2 = \sigma_2^2$

i.e., there is equality of variances of wage distribution in the two cities.

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

$$S_1^2 = \left(\frac{n_1}{n_1-1}\right) s_1^2 = \frac{22}{21} \times (2.9)^2 = 8.81$$

$$S_2^2 = \left(\frac{n_2}{n_2-1}\right) s_2^2 = \frac{16}{15} \times (3.8)^2 = 15.40$$

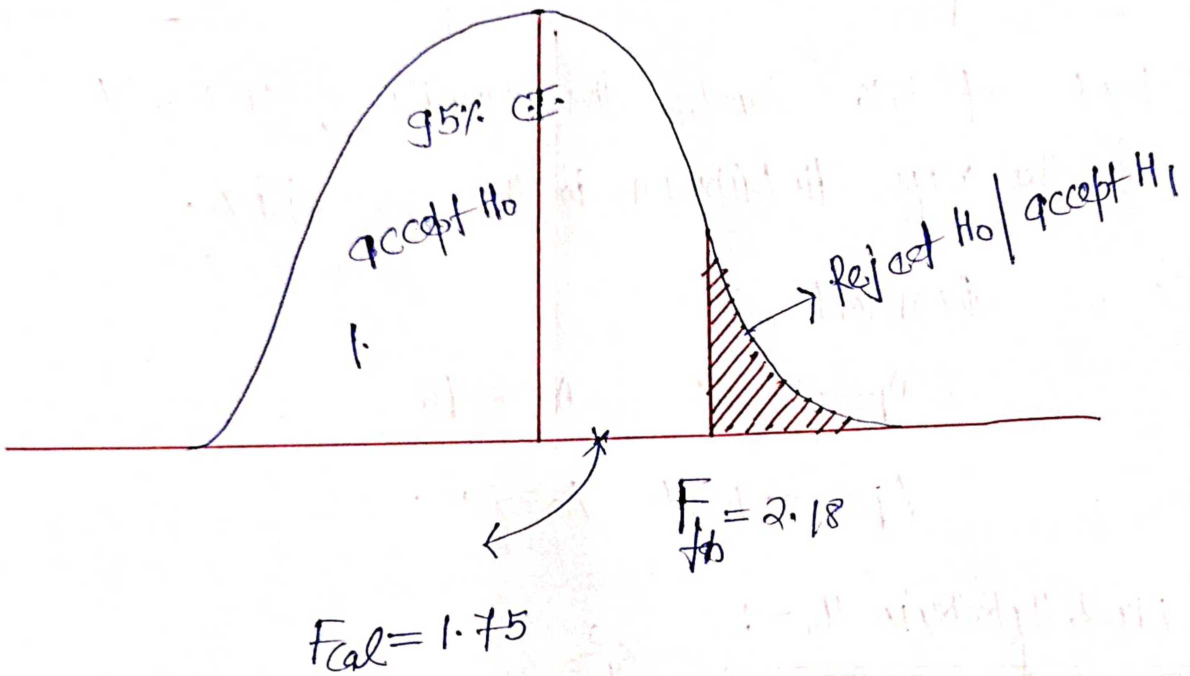
F test

$$F = \frac{S_2^2}{S_1^2} = \frac{15.40}{8.81}$$

$$\boxed{F_{cal} = 1.75}$$

d.f. $= (n_2-1, n_1-1) = (15, 21)$, the tabulated value at 5% level of significance

$$F_{0.05}(15, 21) = 2.18$$



As $F_{cal} = 1.75 < F_{tab} = 2.18$, i.e., Ho is accepted,

i.e., there is equality of variances of the wages distribution in the two cities.

#