Chapter(12)

Analysis of Variance

Test of hypothesis about the difference between Two population Variances

Step (1): State the null (H_0) and alternate (H_1) hypothesis

Case 1: $H_0 : \sigma_1^2 \le \sigma_2^2$
$H_1 : \sigma_1^2 > \sigma_2^2$

Case 2: $H_0 : \sigma_1^2 \ge \sigma_2^2$
$H_1 : \sigma_1^2 < \sigma_2^2$

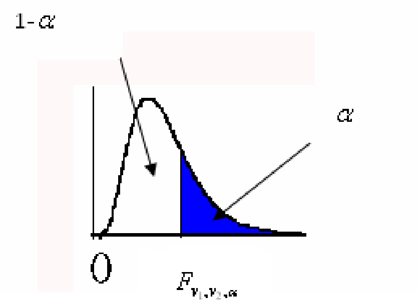
Case 3: $H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 \ne \sigma_2^2$

Step (2): Select a level of significance.
Step (3): Select the Test Statistic (computed value)

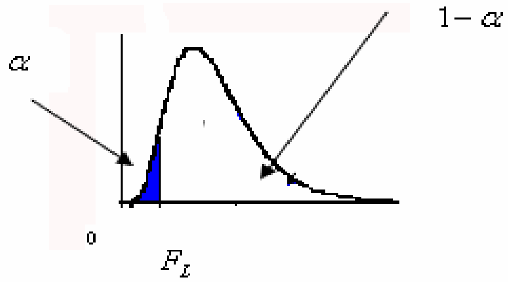
$$F = \frac{S_1^2}{S_2^2}$$

Step (4): Selected the Critical value

The one – tailed test (Right)	$F_{(\nu_1 \nu_2, \alpha)}$

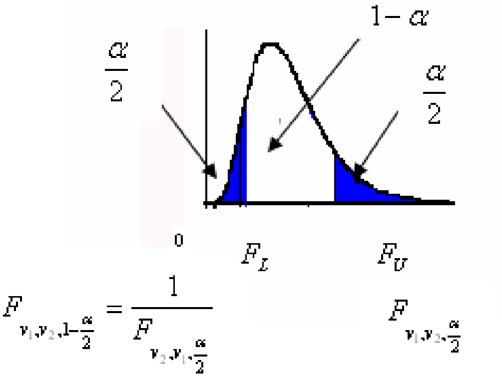


The one – tailed test (left)	$F_{(\nu_1 \nu_2, 1-\alpha)} = \dfrac{1}{F_{(\nu_2 \nu_1, \alpha)}}$



$$F_{\nu_1,\nu_2,1-\alpha} = \frac{1}{F_{\nu_2,\nu_1,\alpha}}$$

The two – tailed test	$F_{\left(\nu_1\nu_2,1-\frac{\alpha}{2}\right)} = \frac{1}{F_{\left(\nu_2\nu_1,\frac{\alpha}{2}\right)}}$ & $F_{\left(\nu_1\nu_2,\frac{\alpha}{2}\right)}$



Step (5): Formulate the Decision Rule and Make a Decision

Case1: Reject H_0 if $F_c > F_{(\nu_1\nu_2,\alpha)}$

Case2: Reject H_0 if $\boldsymbol{F_c} < \frac{\mathbf{1}}{\boldsymbol{F}_{(\nu_2\nu_1,\alpha)}}$

Case3: Reject H_0 if

$F_c > F_{\left(\nu_1\nu_2,\frac{\alpha}{2}\right)}$ or $F_c < \frac{1}{F_{\left(\nu_2\nu_1,\frac{\alpha}{2}\right)}}$

Example (6)

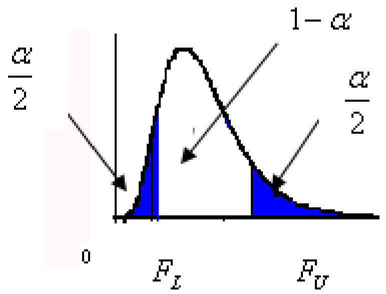
Lammers Limos offers limousine service from the city hall in Toledo, Ohio, to Metro Airport in Detroit. Sean Lammers, president of the company, is considering two routes. One is via U.S. 25 and the other via I-75. He wants to study the time it takes to drive to the airport using each route and then compare the results. He collected the following sample data, which is reported in minutes.

Route	Sample mean	Sample Standard deviations	Sample size
Route 25	58.29 minutes	8.9947 minutes	7
Route i-75	59minutes	4.3753 minutes	8

Using the 0.10 significance level, is there a difference in the variation in the driving times for the two routes?

Solution:

Step 1: State the null hypothesis and the alternate hypothesis.



$H_0 : \sigma_1{}^2 = \sigma_2^2$

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

This is two-tailed test

(Note: keyword in the problem "is there a difference")

Step 2: Select the level of significance.

α = 0.10 as stated in the problem , α/2 = 0.05

Step 3: Select the test statistic.

$$F = \frac{S_1^2}{S_2^2} = \frac{(8.9947)^2}{(4.3753)^2} = 4.23$$

Step 4: Formulate the decision rule.

$$F_{\left(\nu_1 \nu_2, \frac{\alpha}{2}\right)} = F_{(6,7\,,0.05)} = 3.87$$

$$F_{\left(\nu_1 \nu_2, 1-\frac{\alpha}{2}\right)} = \frac{1}{F_{\left(\nu_2 \nu_1, \frac{\alpha}{2}\right)}} = \frac{1}{F_{(7,6\,,0.05)}} = \frac{1}{4.21} = 0.24$$

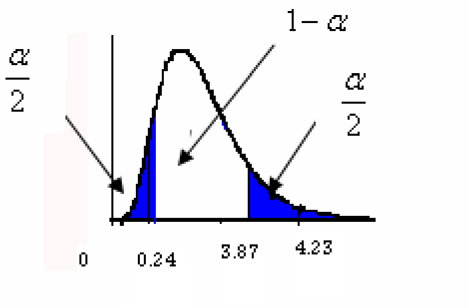
F Table for alpha=.05											
V2	V1										
	1	2	3	4	5	6	7	8	9	10	12
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91

Step 5: Make a decision and interpret the result.

$$\text{Reject } \mathrm{H}_0 \text{ if } \quad F_c > F_{\frac{\alpha}{2}, V_1, V_2} \quad , \quad F_c > F_{0.05,6,7} = 3.86$$

Or

$$F_c < F_{1-\frac{\alpha}{2}, V_1, V_2} = \frac{1}{F_{\frac{\alpha}{2}, \nu_2, \nu_1}} = 0.24$$



The decision is to reject the null hypothesis, because the computed F-Value (4.23) is larger than the critical value (3.87). We conclude that there is a difference in the variation of the travel times along the two routes.

Comparing Means of Two or More Populations
Analysis of Variance (ANOVA)
one way ANOVA

Data frame:

Individuals (Observations)	Groups (Populations)				
	1	2	…. I ……	K	Total
1	Y_{11}	Y_{21}	……. Y_{i1}	Y_{k1}	
2	Y_{12}	Y_{22}	……..Y_{i2}	Y_{k2}	
….	…	….	….	….	
j	Y_{1j}	Y_{2j}	……. Y_{ij}	Y_{kj}	
n	Y_{1n}	Y_{2n}	……..Y_{in}	Y_{kn}	
Sum	$Y_{1.}$	$Y_{2.}$	……..$Y_{i.}$	$Y_{k.}$	$Y_{..}$
Mean	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$	$\bar{Y}_{3.}$	$\bar{Y}_{k.}$	$\bar{Y}_{...}$

Y_{ij} : The jth sample observation selected from group or population i

n_i : The number of sample observations selected from population i

n : The total sample size ($n = n_1 + n_2 + \ldots\ldots + n_k$)

$Y_{i.}$: The sum (total) of the sample measurements obtained from population i.

$Y_{..}$: The grand total.

Step (1): State the null and alternate hypotheses:

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots\ldots\ldots = \mu$

H_1: At least one of the K- population means different from the rest.

Step (2): Select the level of significance (α)
Step (3): The test statistic

Because we are comparing means of more than two groups, use the F statistic.

The F test statistic is found by dividing the between group variance (MSB) by the within group variance (MSW).

$$F_c = \frac{\mathbf{MSB}}{\mathbf{MSW}} = \frac{\mathbf{SSB/K-1}}{\mathbf{SSW/n-K}}$$

SSB: **Sum of Squares Between groups**. (This variation due to the interaction between the samples)

SSW: Sum of Squares Within groups. (This variation due to differences within individual samples).

SST: Total Sum of Squares. (The total variation is comprised the sum of the squares of the differences of each mean with the grand mean)

$$SSB = \sum_{i=1}^{K} n_i\,(\bar{Y}_i - \bar{Y}_{..})^2 = \sum_{i=1}^{k}\left(\frac{Y_{i.}^2}{n_i}\right) - \frac{Y_{..}^2}{n}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 = \sum\sum Y_{ij}^2 - \frac{Y_{..}^2}{n}$$

SSW = SST – SSB

Step (4): The critical value:

The degrees of freedom for the **numerator** are the degrees of freedom for the between group (k-1) and the degrees of freedom for the **denominator** are the degrees of freedom for the within group (n-k).

$$\mathbf{F}_{(\alpha\ ,\ \mathbf{K-1}\ ,\mathbf{n-k}\ )}$$

Step (5) : Formulate the decision Rule and make a decision

Reject H_o

If $\mathbf{F_c} > \mathbf{F}_{(\alpha\ ,\mathbf{K-1},\mathbf{n-K})}$

It is convenient to summarize the calculation of the F statistic in (ANOVA Table)

ANOVA TABLE

Source of variation (S.V)	Sum of Squares (S.S)	Degrees of freedom	Mean Squares (MS)	F- ratio
Between groups	SSB	K-1	MSB =SSB/K-1	F = MSB / MSW
Within groups	SSW	n-K	MSW = SSW/n-k	
Total	SST	n-1		

Example (1):

Recently a group of four major carriers joined in hiring Brunner Marketing Research, Inc., to survey recent passengers regarding their level of satisfaction with a recent flight.

The survey included questions on ticketing, boarding, in-flight service, baggage handling, pilot communication, and so forth. Twenty-five questions offered a range of possible answers: excellent, good, fair, or poor. A response of excellent was given a score of 4, good a 3, fair a 2, and poor a 1. These responses were then totaled, so the total score was an indication of the satisfaction with the flight. Brunner Marketing Research, Inc., randomly selected and surveyed passengers from the four airlines. Is there a difference in the mean satisfaction level among the four airlines? (Construct the ANOVA table)

Use the .01 significance level.

Eastern	TWA	Allegheny	Ozark
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

Solution:

Step 1: State the null and alternate hypotheses.

H_0: $\mu_E = \mu_A = \mu_T = \mu_O$

H1: The means are not all equal

Step 2: State the level of significance.

The .01 significance level is stated in the problem.

Step (3): Find the appropriate test statistic.

$$\mathbf{F_c = \frac{MSB}{MSW} = \frac{SSB/K-1}{SSW/n-K}}$$

Individuals (Observations)	Groups (Populations)				
	E	T	A	O	Total
1	94	75	70	68	
2	90	68	73	70	
3	85	77	76	72	
4	80	83	78	65	
5		88	80	74	
6			68	65	
7			65		
Sum ($Y_{i.}$)	$Y_{1.}$= 349	$Y_{2.}$=391	$Y_{3.}$=510	$Y_{2.}$=414	ΣY_{ij} = Y..= 1664
n	4	5	7	6	22
Mean	$\bar{Y}_{1.}$**=87.25**	$\bar{Y}_{2.}$**=78.2**	$\bar{Y}_{3.}$**=72.86**	$\bar{Y}_{4.}$**=6969**	

$$SST = \sum\sum Y_{ij}^2 - \frac{Y_{..}^2}{n} = (94^2 + 90^2 + \cdots + 65^2) - \left(\frac{1664^2}{22}\right) = 1485.09$$

$$\mathrm{SSB} = \sum_{i=1}^{k}\left(\frac{Y_{i.}^2}{n_i}\right) - \frac{Y_{..}^2}{n} = \left(\frac{349^2}{4} + \frac{391^2}{5} + \frac{510^2}{7} + \frac{414^2}{6}\right) - \left(\frac{1664^2}{22}\right)$$

$$= 890.68$$

SSW = SST – SSB = 1485.09 -890.68 =594.41

$$\mathrm{F} = \frac{\mathrm{MSB}}{\mathrm{MSW}} = \frac{\mathrm{SSB/K-1}}{\mathrm{SSW/n-K}} = \frac{890.68/4-1}{594.41/22-4} = 8.99$$

Step (4): The critical value:

The degrees of freedom for the **numerator = K-1 =4-1=3**

The degrees of freedom for the **denominator = n-K =22-4 =18**

$$\mathbf{F}_{(\alpha\ ,\ \mathbf{K-1},\mathbf{n-k})} = \mathbf{F}_{(0.01\ ,\ 3,18)} = \mathbf{5.09}$$

Step (5): State the decision rule.

$$\mathbf{F_c} > \mathbf{F}_{(0.01,3,18)}$$

Reject H_0

The computed value of F is 8.99, which is greater than the critical value of 5.09, so the null hypothesis is rejected.

Conclusion: The population means are not all equal. The mean scores are not the same for the four airlines; at this point we can only conclude there is a difference in the treatment means. We cannot determine which treatment groups differ or how many treatment groups differ.

ANOVA TABLE

(S.V)	(S.S)	DF	(MS)	F- ratio
Between groups	SSB = 890.68	K-1 = 4-1 =3	MSB =SSR/K-1= 890.86/3 = 296.95	F = MSB / MSW = 296.95/33.02 = 8.99
Within groups	SSW = 594.41	n-K =22-4 =18	MSW = SSE/n-k = 594.41/18 = 33.02	
Total	SST =1485.09	n-1 = 22-1=21		