



Statistical Inference Theory

Lesson 42

Analysis of Variance (ANOVA)

639

Analysis of Variance (ANOVA) is a method for testing the hypothesis of the differences between three or more population means. In the case where we are testing three population means, H_0 and H_a are

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

$$H_a: \mu_1 \neq \mu_2 \text{ or}$$

$$\mu_1 \neq \mu_3 \text{ or}$$

$$\mu_2 \neq \mu_3$$

To use ANOVA, we need to make the following assumptions about the given populations:

1. The populations are assumed to be (approximately) normally distributed.
2. The populations have equal variances.
3. The samples drawn from each population are independent of each other.

The following example is as typical problem to be solved by ANOVA:

Assume the track coach of a local high school wishes to test, among three brands of running shoes, the best performing shoes. He decides to select 15 track runners to run the 100 yard dash. In this race, each brand is worn by five runners. The following table gives the timing outcome of the race for each brand:

Brand A(seconds)	Brand B (seconds)	Brand C (seconds)
11.30	12.17	11.43
12.06	12.14	10.11
11.78	11.66	11.66
12.11	10.99	10.28
10.98	10.89	11.55

42.1 - One-factor classification

Problems of this type are called one-factor classification because only one variable (factor) is considered: the brand of running shoes. The variables are called treatments. For the above example we have three treatments, the three brands of shoes. In this lesson we will study both one-factor and two-factor classification of ANOVA where the samples are of equal size for each treatment.

In deciding to reject H_0 or not, we first must learn to compute three types of variations:

1. Total variation (S_T^2)
2. Variation within treatments (S_W^2)
3. Variation between treatments (S_B^2)

The equation for the relationship between total variation, within variation and between variation is

$$S_W^2 = S_T^2 - S_B^2 .$$

42.1 - Example 1: For the example above, compute

- (a). total variation.
- (b). variation between treatments.
- (c). variation within treatments.

Solutions:

► (a).

Step 1: Compute the mean \bar{X} for all the numbers in the table:

Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
11.30	12.17	11.43
12.06	12.14	10.11
11.78	11.66	11.66
12.11	10.99	10.28
10.98	10.89	11.55
$\bar{X} = 11.41$		

Step 2: • Subtract \bar{X} from each of the numbers in the table.

- Square each of these.
- The total variation is the sum of the values computed in the table:

Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
$(11.30 - 11.41)^2 \approx 0.01$	$(12.17 - 11.41)^2 \approx 0.58$	$(11.43 - 11.41)^2 \approx 0.0$
$(12.06 - 11.41)^2 \approx 0.43$	$(12.14 - 11.41)^2 \approx 0.54$	$(10.11 - 11.41)^2 \approx 1.68$
$(11.78 - 11.41)^2 \approx 0.14$	$(11.66 - 11.41)^2 \approx 0.06$	$(11.66 - 11.41)^2 \approx 0.06$
$(12.11 - 11.41)^2 \approx 0.49$	$(10.99 - 11.41)^2 \approx 0.17$	$(10.28 - 11.41)^2 \approx 1.27$
$(10.98 - 11.41)^2 \approx 0.18$	$(10.89 - 11.41)^2 \approx 0.27$	$(11.55 - 11.41)^2 \approx 0.02$

Sum of Table Values: $S_T^2 \approx 5.92$

►(b).

Step 1: Compute the mean \bar{X} for each column:

Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
11.30	12.17	11.43
12.06	12.14	10.11
11.78	11.66	11.66
12.11	10.99	10.28
10.98	10.89	11.55
$\bar{X} = 11.65$	$\bar{X} = 11.57$	$\bar{X} = 11.00$

Step 2: • Subtract \bar{X} (computed in step 1) from each of the \bar{X} s' in the above table.

- Square each of these differences.
- The between variation is the sum of the values from b multiplied by the number of rows.

$(11.65-11.41)^2 \approx 0.06$
$(11.57-11.41)^2 \approx 0.03$
$(11.01-11.41)^2 \approx 0.16$
$S_B^2 \approx 5x(\text{column sum}) = 5(0.25) = 1.25$

►(c).

To compute the within variation, we use the formula:

$$S_W^2 = S_T^2 - S_B^2 = 5.92 - 1.25 = 4.67.$$

Solved Problems

42.1 - Solved Problem 1: A large petroleum company wishes to test five new gasoline additives for increase fuel efficiency. Their research department purchased 35 new model sedans and drove each car 100 miles, over the same track. Each additive was mixed with the gasoline of seven sedans.

The following table is the mileage recorded for each car in this test. Here, mileage is measured for each car as to the number of gallons consumed to travel 100 miles.

Additive	A	B	C	D	E
Number of gallons	5.56	4.97	5.01	4.11	6.87
	5.87	6.00	4.25	5.11	5.76
	5.00	6.77	5.98	5.12	6.02
	5.73	6.09	7.11	6.11	7.00
	3.99	4.78	5.55	6.00	6.51
	4.12	5.00	5.44	5.51	6.09
	4.78	5.11	5.54	7.00	7.11

Compute:

- (a). total variation.
- (b). variation between treatments.
- (c). variation within treatments.

Solutions:

►(a).

Step 1: Compute the mean \bar{X} for all the numbers in the table:

A	B	C	D	E
5.56	4.97	5.01	4.11	6.87
5.87	6.00	4.25	5.11	5.76
5.00	6.77	5.98	5.12	6.02

5.73	6.09	7.11	6.11	7.00
3.99	4.78	5.55	6.00	6.51
4.12	5.00	5.44	5.51	6.09
4.78	5.11	5.54	7.00	7.11

$\bar{X} = 5.63$

Step 2: • Subtract \bar{X} from each of the numbers in the table.

- Square each of these.
- The total variation is the sum of the values in the table:

A	B	C	D	E
$(5.56 - 5.63)^2 \approx 0.00$	$(4.97 - 5.63)^2 \approx 0.43$	$(5.01 - 5.63)^2 \approx 0.38$	$(4.11 - 5.63)^2 \approx 2.30$	$(6.87 - 5.63)^2 \approx 1.54$
$(5.87 - 5.63)^2 \approx 0.06$	$(6.00 - 5.63)^2 \approx 0.14$	$(4.25 - 5.63)^2 \approx 1.90$	$(5.11 - 5.63)^2 \approx 0.27$	$(5.76 - 5.63)^2 \approx 0.01$
$(5.00 - 5.63)^2 \approx 0.39$	$(6.77 - 5.63)^2 \approx 1.30$	$(5.98 - 5.63)^2 \approx 0.12$	$(5.12 - 5.63)^2 \approx 0.26$	$(6.02 - 5.63)^2 \approx 0.15$
$(5.73 - 5.63)^2 \approx 0.01$	$(6.09 - 5.63)^2 \approx 0.21$	$(7.11 - 5.63)^2 \approx 2.20$	$(6.11 - 5.63)^2 \approx 0.23$	$(7.00 - 5.63)^2 \approx 1.88$
$(3.99 - 5.63)^2 \approx 2.68$	$(4.78 - 5.63)^2 \approx 0.72$	$(5.55 - 5.63)^2 \approx 0.00$	$(6.00 - 5.63)^2 \approx 0.14$	$(6.51 - 5.63)^2 \approx 0.78$
$(4.12 - 5.63)^2 \approx 2.27$	$(5.00 - 5.63)^2 \approx 0.39$	$(5.44 - 5.63)^2 \approx 0.04$	$(5.51 - 5.63)^2 \approx 0.01$	$(6.09 - 5.63)^2 \approx 0.21$
$(4.78 - 5.63)^2 \approx 0.72$	$(5.11 - 5.63)^2 \approx 0.27$	$(5.54 - 5.63)^2 \approx 0.01$	$(7.00 - 5.63)^2 \approx 1.88$	$(7.11 - 5.63)^2 \approx 2.20$

Sum of the table values: $S^2_T \approx 26.15$

►(b).

Step 1: Compute the mean \bar{X} for each column:

A	B	C	D	E
5.56	4.97	5.01	4.11	6.87
5.87	6.00	4.25	5.11	5.76
5.00	6.77	5.98	5.12	6.02
5.73	6.09	7.11	6.11	7.00
3.99	4.78	5.55	6.00	6.51
4.12	5.00	5.44	5.51	6.09
4.78	5.11	5.54	7.00	7.11
$\bar{X} \approx 5.01$	$\bar{X} \approx 5.53$	$\bar{X} \approx 5.55$	$\bar{X} \approx 5.57$	$\bar{X} \approx 6.48$

Step 2: • Subtract \bar{X} (computed in step 1) from each of the \bar{X}_s' in the above table.

- Square each of these differences.
- The between variation is the sum of the values from b multiplied by the number of rows.

$(5.01 - 5.63)^2 \approx 0.39$
$(5.53 - 5.63)^2 \approx 0.01$
$(5.55 - 5.63)^2 \approx 0.01$
$(5.57 - 5.63)^2 \approx 0.00$
$(6.48 - 5.63)^2 \approx 0.73$
$S_B^2 \approx 7 \times (\text{table Sum}) = 7(1.13) = 7.91$

(C). To compute the within variation, we use the formula:

$$S_W^2 = S_T^2 - S_B^2 = 26.15 - 7.91 = 18.24 .$$

Unsolved Problems with Answers

42.1 - Problem 1: A medical research laboratory wishes to test if there is a difference between three different drugs that promote weight loss for women over 200 pounds. The client randomly divide up fifteen over-weight women into three equal groups. Each group takes only one of the drugs. The following table is the resulting weight loss (in pounds) after 60 days:

Drug A	Drug B	Drug C
21.30	22.11	33.43
32.06	19.18	29.00
21.78	17.66	21.26
22.11	19.79	30.58
30.98	20.89	21.55

Compute:

- total variation.
- variation between treatments.
- variation within treatments.

Answers:

►(a). $S_T^2 \approx 394$

►(b). $S_B^2 \approx 145$

►(c). $S_W^2 \approx 248$

↑↑ Refer back to 42.1 - Example 1 & 42.1 - Solved Problem 1.

42.2 - Testing Hypothesis on Means using the F Distribution

To test the null and alternative hypothesis:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$$

$$H_a: \mu_1 \neq \mu_2$$

or

$$\mu_1 \neq \mu_3$$

or

$$\mu_2 \neq \mu_3, \text{ etc,}$$

we use the F distribution.

Using the values from analysis of variance, we need to test the F distribution for the value

$$F = \frac{S_B^2(c-1)}{S_W^2(c-1)}, \text{ where}$$

c = the number of treatments (number of columns of the tables)

r = the sample size for each treatment (number of rows of the tables)

$d_2 = c - 1$ degrees of freedom

$d_1 = c(r - 1)$ degrees of freedom.

42.2 - Example 1: For the 42.1- Example 1,

(a). State H_0 and H_a .

(b). Compute F.

(c). Find $F_{0.05}$.

Would you reject H_0 ?

State your conclusions.

(d). Find $F_{0.01}$.

Would you reject H_0 ?

State your conclusions.

Solutions:

►(a).

$$H_0: \mu_A = \mu_B = \mu_C$$

H_a : at least one of the μ values is different from the other two.

►(b).

From 42.1 - Example 1,

$$S_B^2 = 1.25,$$

$$S_W^2 = 4.67,$$

$c = 3$ and $r = 5$,

$$F = \frac{S_B^2 c(r-1)}{S_W^2 (c-1)} = \frac{(1.25)3(5-1)}{(4.67)(3-1)} = \frac{15}{9.34} \approx 1.61.$$

►(c).

$$d_2 = c - 1 = 3 - 1 = 2$$

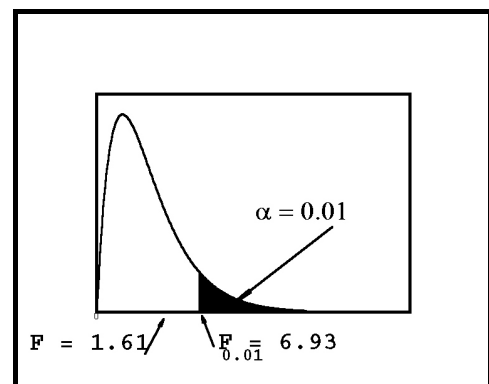
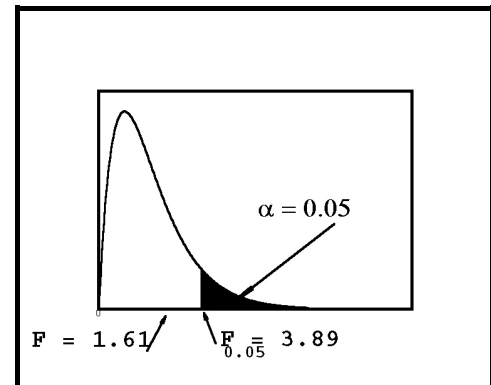
$$d_1 = c(r - 1) = 3(5 - 1) = 12$$

From the F distribution table for 0.05, we find $F_{0.05} = 3.89$.

Since $F = 1.61 < 3.89$, H_0 is not rejected. For a level of significance of 0.05 we have not statistical basis to conclude that the make of the running shoes improves the performance of the runners.

►(d).

$$d_2 = c - 1 = 3 - 1 = 2$$



$$d_1 = c(r - 1) = 3(5 - 1) = 12$$

From the F distribution table for 0.01, we find $F_{0.01} = 6.93$. Since $F = 1.61 < 6.93$, H_0 is not rejected. For a level of significance of 0.01, we have not statistical basis to conclude that the make of the running shoes improves the performance of the runners.

Solved Problems

42.2 Solved Problem 1: For solved 42.1 - Problem 1,

(a). State H_0 and H_a .

(b). Compute F.

(c). Find $F_{0.05}$.

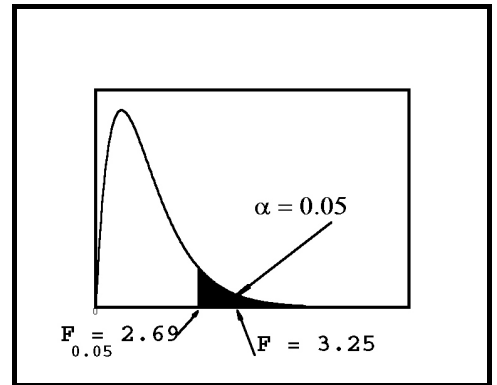
Would you reject H_0 ?

State your conclusions.

(d). Find $F_{0.01}$.

Would you reject H_0 ?

State your conclusions.



Solutions:

►(a).

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$$

H_a : at least one of the μ values is different from the other two.

►(b).

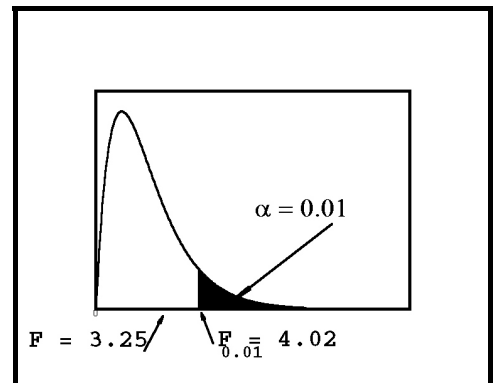
From 42.1 - Solved Problem 1,

$$S_B^2 = 7.91,$$

$$S_W^2 \approx 18.24,$$

$c = 5$ and $r = 7$,

$$F = \frac{S_B^2 c(r-1)}{S_W^2 (c-1)} = \frac{(7.91)5(7 - 1)}{(18.24)(5-1)} = \frac{237.3}{72.96} \approx 3.25 .$$



►(c).

$$d_2 = c - 1 = 5 - 1 = 4$$

$$d_1 = c(r - 1) = 5(7 - 1) = 30$$

From the F distribution table for 0.05, we find $F_{0.05} = 2.69$.

Since $F = 3.25 > 2.69$, H_0 is rejected. For a level of significance of 0.05 we have a statistical basis to conclude that the type of additive makes a difference in mileage.

►(d).

$$d_2 = c - 1 = 5 - 1 = 4$$

$$d_1 = c(r - 1) = 5(7 - 1) = 30$$

From the F distribution table for 0.01, we find $F_{0.01} = 4.02$.

Since $F = 3.25 < 4.02$, H_0 is not rejected. For a level of significance of 0.01, we have no statistical basis to conclude that the type of additive makes a difference in mileage.

Unsolved Problems with Answers

42.2 - Problem 1: For unsolved 42.1 - Problem 1,

(a). State H_0 and H_a .

(b). Compute F.

(c). Find $F_{0.05}$.

Would you reject H_0 ?

State your conclusions.

(d). Find $F_{0.01}$.

Would you reject H_0 ?

State your conclusions.

Answers:

►(a). $H_0: \mu_A = \mu_B = \mu_C$

H_a : at least one of the μ values is different from the other two.

►(b). $F \approx 3.49$

►(c). $F_{0.05} = 3.89$

Since $F = 3.49 < 3.89$, we do not reject H_0 . There is no statistical basis for assuming there is any difference among the three diet drugs for reducing weight.

►(d). $F_{0.01} = 6.93$

Since $F = 3.49 < 6.93$, we do not reject H_0 . There is statistical basis for assuming there is any difference among the three diet drugs for reducing weight.

↑↑ Refer back to 42.2 - Example 1 & 42.2 - Solved Problem 1.

42.3 - Two-factor classification

For two factor classification, we have two different types of treatments. In table form, one set of treatments will be list in the top row and the second set of treatments will be listed on the first column. The following is an example of a two - factor classification of analysis of variance.

Assume the track coach of a local high school wishes to test, among three brands of running shoes, the best performing shoes for freshmen, sophomore, and senior students. The numbers in the table below, represent the average running times for the 100 yard dash.

	Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
Freshmen	10.80	10.27	10.63
Sophomore	11.66	11.84	11.11
Senior	10.71	12.06	12.06

Since we have two different types of treatments, we have two null hypothesis to tests:

$H_0^{(1)}$: There is no difference between brands of running shoes (between columns).

$H_0^{(2)}$: There is no difference between class of the runners. (between rows).

In deciding to reject $H_0^{(1)}$ or $H_0^{(2)}$, we first must learn to compute four types of variations:

1. Total variation (S_T^2)
2. Variation between rows (S_R^2)
3. Variation between columns (S_C^2)
4. Variation due to chance (S_E^2)

The equation for the relationship between these four variations is $S_E^2 = S_T^2 - S_R^2 - S_C^2$.

42.3 - Example 1: For the example above, compute:

- (a). total variation.
- (b). variation between rows.
- (c). variation between columns.
- (d). random variation.

Solutions:

►(a).

Step 1: From the table above, compute the row totals, column totals, row means, column means, table total and mean of the table:

	Brand A (seconds)	Brand B (seconds)	Brand C (seconds)	Row total	Row Mean
Freshmen	10.80	10.27	10.63	31.7	10.57
Sophomore	11.66	11.84	11.11	34.61	11.54
Senior	10.71	12.06	12.06	34.83	11.61
Column Total	33.17	34.17	33.80	Table Total	Table Mean
Column Mean	11.06	11.39	11.27		

Step 2: Complete the following table by subtracting the grand mean from each value of the table and squaring these differences:

	Brand A (seconds)	Brand B (seconds)	Brand C (seconds)
Freshmen	$(10.80 - 11.24)^2 \approx 0.19$	$(10.27 - 11.24)^2 \approx 0.94$	$(10.63 - 11.24)^2 \approx 0.37$
Sophomore	$(11.66 - 11.24)^2 \approx 0.18$	$(11.84 - 11.24)^2 \approx 0.36$	$(11.11 - 11.24)^2 \approx 0.02$
Senior	$(10.71 - 11.24)^2 \approx 0.28$	$(12.06 - 11.24)^2 \approx 0.67$	$(12.06 - 11.24)^2 \approx 0.68$

Step 3: Total variation is the total of all these numbers:

$$S_T^2 \approx 3.69.$$

►(b).

The formula for the variation between rows is by summing the values in the following table:

(Row Mean - Table Mean)²
$(10.57 - 11.24)^2 \approx 0.45$
$(11.54 - 11.24)^2 \approx 0.09$
$(11.61 - 11.24)^2 \approx 0.14$
Sum = 0.68
$S_R^2 = c\text{Sum} = 3(0.68) = 2.04$

►(c).

The formula for the variation between columns is by summing the values in the following table:

(Column mean - Table Mean)²
$(11.06 - 11.24)^2 \approx 0.032$
$(11.39 - 11.24)^2 \approx 0.023$
$(11.27 - 11.24)^2 \approx 0.00$
Sum = 0.055
$S_C^2 = r\text{Sum} = 3(0.055) = 0.165$

►(d).

To compute the random variation (S_E^2), we use the formula:

$$S_E^2 = S_T^2 - S_R^2 - S_C^2 = 3.69 - 2.04 - 0.165 \approx 1.49$$

Solved Problems

42.3 - Solved Problem 1: A large petroleum company wishes to test five new gasoline additives for increased fuel efficiency. Their research department purchased 15 new model sedans and drove each car 100 miles, over the same track. Each additive was mixed with the three octane gasolines: regular, premium and super. The following table is the mileage recorded for each car in this test. Here, mileage is measured for each car as to the number of gallons consumed to travel 100 miles.

Additive \ Octane	A	B	C	D	E
Regular	5.11	5.23	6.13	5.00	6.18
Premium	4.76	5.00	4.02	5.11	4.87
Super	4.01	4.5	3.98	4.98	5.00

From this table, compute:

- (a). total variation.
- (b). variation between rows.
- (c). variation between columns.
- (d). random variation.

Solutions:

►(a).

Step 1: From the table above, compute the row totals, column totals, row means, column means, grand total and mean of the grand total:

Additive \ Octane	A	B	C	D	E	Row Total	Row Mean
Regular	5.11	5.23	6.13	5.00	6.18	27.65	5.53
Premium	4.76	5.00	4.02	5.11	4.87	23.76	4.75
Super	4.01	4.5	3.98	4.98	5.00	22.47	4.49
Column Total	13.88	14.73	14.13	15.09	16.05	Table Total	Table Mean
Column Mean	4.63	4.91	4.71	5.03	5.35	73.88	4.93

Step 2: Complete the following table by subtracting the table mean value of the and squaring these differences:

Additive \ Octane	A	B	C	D	E
Regular	$(5.11 - 4.93)^2$ ≈ 0.03	$(5.23 - 4.93)^2$ ≈ 0.09	$(6.13 - 4.93)^2$ ≈ 1.44	$(5.00 - 4.93)^2$ ≈ 0.01	$(6.18 - 4.93)^2$ ≈ 1.57
Premium	$(4.76 - 4.93)^2$ ≈ 0.03	$(5.00 - 4.93)^2$ ≈ 0.01	$(4.02 - 4.93)^2$ ≈ 0.82	$(5.11 - 4.93)^2$ ≈ 0.03	$(4.87 - 4.93)^2$ ≈ 0.00
Super	$(4.01 - 4.93)^2$ ≈ 0.84	$(4.45 - 4.93)^2$ ≈ 0.18	$(3.98 - 4.93)^2$ ≈ 0.89	$(4.98 - 4.93)^2$ ≈ 0.00	$(5.00 - 4.93)^2$ ≈ 0.01

Total variation is the total of all these numbers:

$$S_T^2 \approx 5.97$$

►(b).

The formula for variation between rows is by summing the values in the following table:

(Row Means - Table Mean)²
$(5.53 - 4.93)^2 \approx 0.36$
$(4.75 - 4.93)^2 \approx 0.03$
$(4.49 - 4.93)^2 \approx 0.19$
Sum ≈ 0.58
$S_R^2 \approx c(\text{Sum}) = 5(0.58) = 2.90$

►(c).

The formula for variation between columns is by summing the values in the following table:

(Column Means - Table Mean)²
$(4.63 - 4.93)^2 \approx 0.09$
$(4.91 - 4.93)^2 \approx 0.00$
$(4.71 - 4.93)^2 \approx 0.05$
$(5.03 - 4.93)^2 \approx 0.01$
$(5.35 - 4.93)^2 \approx 0.18$
Sum ≈ 0.33
$S_C^2 \approx r(\text{Sum}) = 3(0.33) = 0.99$

►(d).

To compute the random variation (S_E^2), we use the formula:

$$S_E^2 = S_T^2 - S_R^2 - S_C^2 = 5.97 - 2.90 - 0.99 = 2.08$$

Unsolved Problems with Answers

42.3 - Problem 1: A medical research laboratory wishes to test if there is a difference between three different drugs that promote weight loss for women and men over 200 pounds. The following table is the resulting weight loss (in pounds) after 60 days.

Drug/Gender	A	B	C
MALE	22.45	20.65	24.11
FEMALE	27.22	28.00	28.11

From this table, compute

- (a). total variation.
 (b). variation between rows.
 (c). variation between columns.
 (d). random variation.

Answers:

►(a). 49.77

►(b). 43.44

►(c). 3.37

►(d). 2.96

↑↑ Refer back to 42.3 - Example 1 & 42.3 - Solved Problem 1.

42.4 - Testing Hypothesis between rows and between columns using the F Distribution.

We need to test two hypothesis:

$H_0^{(1)}$: There is no statistical significant difference between columns.

$H_0^{(2)}$: There is no statistical significant difference between rows.

To test $H_0^{(1)}$, we use the F distribution where

$$F = \frac{(r - 1)S_C^2}{S_E^2}$$

with $d_2 = c - 1$

and $d_1 = (r - 1)(c - 1)$ degrees of freedom.

To test $H_0^{(2)}$, we use the F distribution where $F = \frac{(c - 1)S_R^2}{S_E^2}$ with $d_2 = r - 1$

and $d_1 = (r - 1)(c - 1)$ degrees of freedom.

42.4 - Example 1: For 42.3 - Example 1,

(a). Find F.

Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between brands of running shoes.

(b). Find F.

Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between class year.

Solutions:

►(a).

Here, we are testing across columns.

Step 1: To find F, we use the formula:

$$F = \frac{(r - 1)S_C^2}{S_E^2},$$

where

r = the number of rows.

Step 2: From Example 3.1, we computed:

$$S_C^2 = 0.15, S_E^2 = 1.56 .$$

Since r = 3,

$$F = \frac{(r - 1)S_C^2}{S_E^2} = \frac{(3 - 1)0.15}{1.56} \approx 0.19 .$$

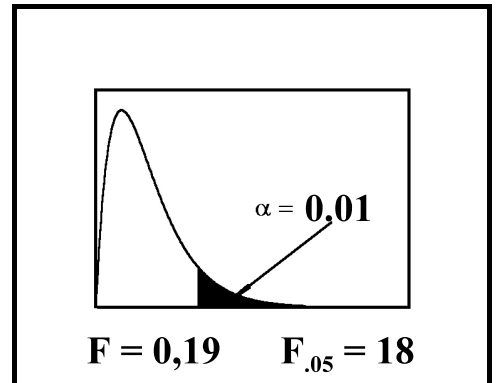
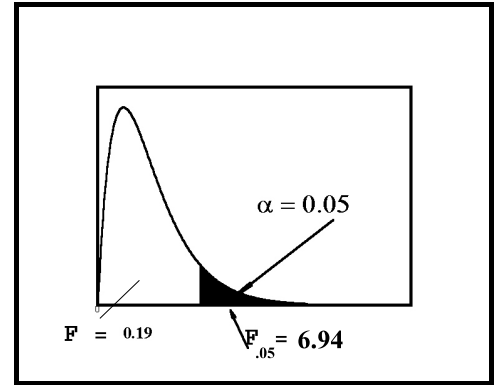
Step 3: We have $d_2 = c - 1 = 3 - 1 = 2$ degrees of freedom and

$d_1 = (r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$ degrees of freedom.

Step 4: Using the F distribution table for $\alpha = 0.05$, $F_{0.05} = 6.94$.

Step 5: Since $F = 0.019 < 6.94$, we conclude there is no significant difference between running shoes.

Step 6: Using the F distribution table for $\alpha = 0.01$, $F_{0.01} = 18$.



Step 7: Since $F = 0.019 < 18$, we conclude there is no significant difference between running shoes.

►(b).

Here, we are testing down rows.

Step 1: To find F , we use the formula:

$$F = \frac{(c - 1)S_R^2}{S_E^2}, \text{ where } c = \text{the number of columns.}$$

Step 2: From Example 3.1, we computed:

$$S_R^2 = 01.98, S_E^2 = 1.56$$

Since $c = 3$,

$$F = \frac{(c - 1)S_R^2}{S_E^2} = \frac{(3 - 1)1.98}{1.56} \approx 2.54 .$$

Step 3: We have $d_2 = r - 1 = 3 - 1 = 2$ degrees of freedom and

$$d_1 = (r - 1)(c - 1) = (3 - 1)(3 - 1) = 4 \text{ degrees of freedom.}$$

Step 4: Using the F distribution table for $\alpha = 0.05$, $F_{0.05} = 6.94$.

Step 5: Since $F = 2.54 < 6.94$, we conclude there is no significant difference between class years.

Step 6: Using the F distribution table for $\alpha = 0.01$, $F_{0.01} = 18$.

Step 7: Since $F = 0.019 < 18$, we conclude there is no significant difference between class years.

Solved Problems

42.4 - Problem 1: For 42.3 - Solved Problem 1,

(a). Find F . Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between gasoline additives.

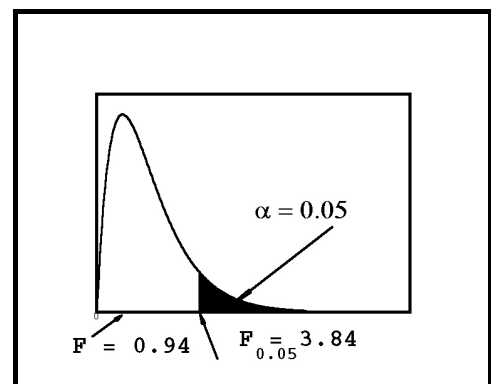
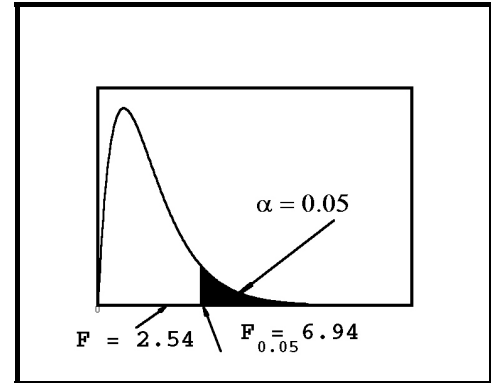
(b). Find F . Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between octanes.

Solutions:

►(a).

Here, we are testing across columns.

Step 1: To find F , we use the formula:



$$F = \frac{(r - 1)S_C^2}{S_E^2} \text{ where } r = \text{the number of rows,}$$

Step 2: From 42.3 - solved problem 1, we computed:

$$S_C^2 = 0.99, \text{ and } S_E^2 = 2.08 .$$

Since $r = 3$,

$$F = \frac{(r - 1)S_C^2}{S_E^2} = \frac{(3 - 1)0.99}{2.08} \approx 0.94 .$$

Step 3: We have $d_2 = 5 - 1 = 5 - 1 = 4$ degrees of freedom and

$$d_1 = (r - 1)(c - 1) = (3 - 1)(5 - 1) = 8 \text{ degrees of freedom.}$$

Step 4: Using the F distribution table for $\alpha = 0.05$, $F_{0.05} = 3.84$.

Step 5: Since $F = 0.95 < 3.84$, we conclude there is no significant difference between gasoline additives.

Step 6: Using the F distribution table for $\alpha = 0.01$, $F_{0.01} = 7.01$.

Step 7: Since $F = 0.95 < 7.08$, we conclude there is no significant difference between gasoline additives.

►(b).

Here, we are testing down rows.

Step 1: To find F, we use the formula:

$$F = \frac{(c - 1)S_R^2}{S_E^2} \text{ where } c = \text{the number of columns.}$$

Step 2: From 42.3 - Solved Problem 1, we computed:

$$S_R^2 = 2.9, \text{ } S_E^2 = 2.08$$

Since $c = 5$,

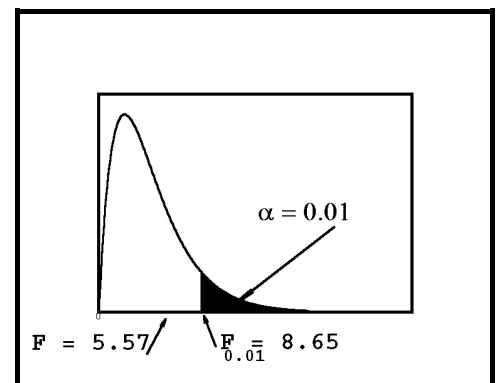
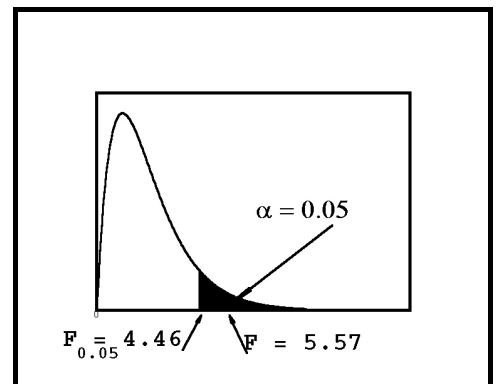
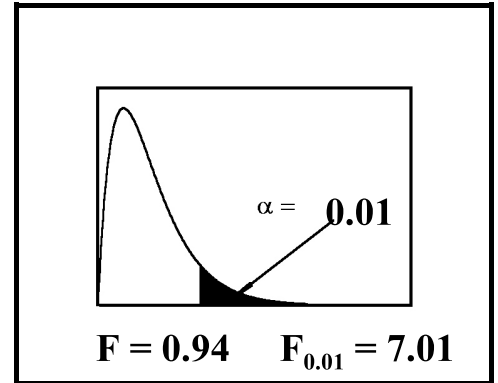
$$F = \frac{(c - 1)S_R^2}{S_E^2} = \frac{(5 - 1)2.90}{2.08} \approx 5.57$$

Step 3: We have $d_2 = r - 1 = 3 - 1 = 2$ degrees of freedom and

$$d_1 = (r - 1)(c - 1) = (3 - 1)(5 - 1) = 8 \text{ degrees of freedom.}$$

Step 4: Using the F distribution table for $\alpha = 0.05$, $F_{0.05} = 4.46$.

Step 5: Since $F = 5.57 > 4.46$, we conclude there is a significant difference between octane which affects



mileage.

Step 6: Using the F distribution table for $\alpha = 0.01$, $F_{0.01} = 8.65$

Step 7: Since $F = 5.57 < 8.65$, we conclude there is no significant difference in octane.

Unsolved Problems with answers.

42.4 - Problem 1: For unsolved 42.3 - Problem 1,

(a). Find F. Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between diet drugs.

(b). Using a level of significance of 0.05 and 0.01, determine if there is a statistical difference between men and women.

Answers:

►(a). Since $F = 1.14$, using a level of significance of 0.05 and 0.01, we conclude there is significant difference between diet drugs.

►(b). $F \approx 29.35$. Since $F = 29.35 > 19$, we conclude there is a significant difference in weight loss between men and women. at a 0.05 level of significance. However, at a 0.01 level of significance, there is no significant weight loss difference between men and women.

↑↑ Refer back to 42.4 - Example 1 & 42.4 - Solved Problem 1.

Supplementary Problems

Ms. Romano teaches three sections of beginning Latin at a local high school. The following table gives the average grades over the past five years.

Latin Section/Average grade	Section 1	Section 2	Section 3
1991	78.20	76.50	79.20
1992	79.10	77.90	79.87
1993	79.60	77.60	78.66
1994	80.10	81.78	79.10
1995	81.10	80.10	85.78

Using one - factor classification to see if there is a significant difference between sections, find:

1. total variation (S_T^2).
2. variation between treatments (S_B^2).
3. Variation within treatments (S_W^2).

4. F.

5. Using a 0.05 level of significance, is there is a difference in grades between class sections.

Applying two factor classification to the above table data, find

6. total variation.

7. variation between rows.

8. variation between columns.

9. random variation.

10. F. Using a 0.05 level of significance, is there is a difference in grades between class sections?

11. F. Using a 0.05 level of significance, is there is a difference in grades between the five years?
