Statistical Inference Theory
Lesson 42
Analysis of Variance(Anova)

42.1 - One-factor classification.

42.1 - Problem 1:

➤(a).

Step 1: For the table compute $\overline{X}$ for the table numbers:

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
$\overline{X} = 24.25$		

Step 2:

- Subtract $\overline{X}$ from each of the numbers in the table.
- Square each of these.
- The total variation is the sum of the values computed in b.

Drug A	Drug B	Drug C
$(21.30 - 24.25)^2 = 8.70$	$(22.11 - 24.25)^2 = 4.58$	$(33.43 - 24.25)^2 = 84.27$
$(32.06 - 24.25)^2 = 61$	$(19.18 - 24.25)^2 = 25.70$	$(29.00 - 24.25)^2 = 22.56$
$(21.78 - 24.25)^2 = 6.1$	$(17.66 - 24.25)^2 = 43.43$	$(21.26 - 24.25)^2 = 8.94$
$(22.11 - 24.25)^2 = 4.58$	$(19.79 - 24.25)^2 = 19.89$	$(30.58 - 24.25)^2 = 40.07$
$(30.98 - 24.25)^2 = 45.30$	$(20.89 - 24.25)^2 = 11.29$	$(21.55 - 24.25)^2 = 71.29$
SUM OF TABLE VALUES:	$S_T^2 \approx 394$	

➤(b).
Step 1: Compute the mean $\overline{X}$ for each column:

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
$\overline{X} \approx 25.65$	$\overline{X} \approx 19.93$	$\overline{X} \approx 27.16$

Step 2:
- Subtract $\overline{X}$ (computed in step 1) from each of the $\overline{X}s'$ 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.

$(25.65 - 24.25)^2 \approx 1.96$
$(19.93 - 24.25)^2 \approx 18.66$
$(27.16 - 24.25)^2 \approx 8.47$
S_B^2 = 5x(column sum) = 5(29.10) $\approx$ 145

➤(c). variation within treatments.

Solution:

To compute the within variation , we use the formula:

$$S_W^2 = S_T^2 - S_B^2 = 394 - 145 = 249$$

42.2 - Testing Hypothesis on Means using the F Distribution.

42.2 - Problem 1: For problem 41.1,

➤(a).

H_o: $\mu_A = \mu_B = \mu_C$

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

➤(b).

From 42.1 - Problem 1,

$S_B^2 = 145,$

$S_W^2 = 249,$

c = 3 and r = 5.

$$F = \frac{S_B^2 c(r-1)}{S_W^2(c-1)} = \frac{(145)3(5-1)}{(249)(3-1)} = \frac{1740}{498} \approx 3.49$$

➤(c).

From the formula for degree of freedom,

$d_2 = c - 1 = 3 - 1 = 2$, where c is the number of columns.

$d_1 = c(r - 1) = 3(5 - 1) = 12$, where r is the number of rows.

From the D table, $F_{0.05} = 3.89$. 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).

From the D table, $F_{0.01} = 6.93$. Since $F = 3.49 < 6.93$, 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.

42.3 - Two-factor classification.

42.3 - Problem 1:
➤(a).

Step 1:

Drug/Gender	A	B	C	Row Total	Row mean
MALE	22.45	20.65	24.11	67.21	22.40
FEMALE	27.22	28.00	28.11	83.33	27.78
Column total	49.67	48.65	52.22	**Total table**	**Table mean**
Column mean	24.84	24.33	26.11	150.54	25.09

Step 2:

Drug/Gender	A	B	C
MALE	$(22.45 - 25.09)^2 \approx 6.97$	$(20.65 - 25.09)^2 \approx 19.71$	$(24.11 - 25.09)^2 \approx 0.96$
FEMALE	$(27.22 - 25.09)^2 \approx 4.54$	$(28.00 - 25.09)^2 \approx 8.47$	$(28.11 - 25.09)^2 \approx 9.12$

The total of the numbers in the above table gives $S_T^2 \approx 49.77$.

➤(b).
The formula for variation between rows is by summing the values in the following table:

(Row Means - Table Mean)^2
$(22.40 - 25.09)^2 \approx 7.24$
$(27.78 - 25.09)^2 \approx 7.24$
Sum ≈ 14.48
$S_R^2 \approx c(\text{Sum}) = 3(14.48) = 43.44$

➤(c).
The formula for variation between columns is by summing the values in the following table:

(Column Means - Table Mean)^2
$(24.84 - 25.09)^2 \approx 0.063$
$(24.33 - 25.09)^2 \approx 0.58$
$(26.11 - 25.09)^2 \approx 1.04$
Sum ≈ 1.68
$S_C^2 \approx 2(\text{Sum}) = 3(1.68) = 3.37$

➤(d).

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

$$S_E^2 = S_T^2 - S_R^2 - S_C^2 = 49.77 - 43.44 - 3.37 = 2.96 .$$

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

42.4 - Problem 1:

➤(a).

Here, we are testing across columns.

Step 1:To find F, we use the formula: $F = \dfrac{(r - 1)S_C^2}{S_E^2}$, where r = the number of rows.

Step 2: From Problem 3.1, we computed: $S_C^2 = 3.37$, and $S_E^2 = 2.96$.

Since r = 3, $F = \dfrac{(r - 1)S_C^2}{S_E^2} = \dfrac{(2 - 1)3.37}{2.96} \approx 1.14$.

Step 3: We have $d_2 = c - 1 = 3 - 1 = 2$ degrees of freedom and $d_1 = (r -1)(c - 1) = (2 -1)(3 -1) = 2$ degrees of freedom.

Step 4: Using the F distribution table for $\alpha = 0.05$, $F_{005} = 19$.

Step 5: Since F = 1.14 < 19, we conclude there is no significant difference between drugs.

Step 6: Using the F distribution table for $\alpha = 0.01$, $F_{0.01} = 99$.

Step 7: Since F = 1.14 < 99, we conclude there is no significant difference between drugs.

➤(b).
Here, we are testing down rows.

Step 1:To find F, we use the formula: $F = \frac{(c - 1)S_R^2}{S_E^2}$, where c = the number of columns.

Step 2: From 42.3 - Example 1, we computed: $S_R^2 = 43.44$, and $S_E^2 = 2.96$.

Since c = 3, $F = \frac{(c - 1)S_R^2}{S_E^2} \quad \frac{(3 - 1)43.44}{2.96} \approx 29.35$.

Step 3: We have $d_2 = r - 1 = 3 - 1 = 2$ degrees of freedom and $d_1 = (r -1)(c - 1) = (2 -1)(3 -1) = 2$ degrees of freedom.

Step 4: Using the F distribution table for $\alpha = 0.05$, $F_{0.05} = 19$.

Step 5: Since F = 29.35 > 19, we conclude there is significant difference between men and women.

Step 6: Using the F distribution table for $\alpha = 0.01$, $F_{001} = 99$.

Step 7: Since F = 29.35 > 99, we conclude there is no significant difference between men and women.

Supplementary Problems

1.

Step 1:

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
$\overline{X} \approx 79.64$			

Step 2:

Section 1	Section 2	Section 3
$(78.20- 79.64)^2 \approx 2.07$	$(76.50 - 79.64)^2 \approx 9.86$	$(79.20 - 79.64)^2 \approx 0.19$
$(79.10 - 79.64)^2 \approx 0.29$	$(77.90- 79.64)^2 \approx 3.03$	$(79.87 - 79.64)^2 \approx 0.053$
$(79.60 - 79.64)^2 \approx 0.002$	$(77.6 - 79.64)^2 \approx 4.16$	$(78.66 - 79.64)^2 \approx 0.96$
$(80.10 - 79.64)^2 \approx 0.21$	$(81.78 - 79.64)^2 \approx 4.58$	$(79.10- 79.64)^2 \approx 0.29$
$(81.10 - 79.64)^2 \approx 2.13$	$(80.10 - 79.64)^2 \approx 0.21$	$(85.78 - 79.64)^2 \approx 37.70$

Summing the values in the above table, equals the total variation

$S_T^2 \approx 65.75$.

2.

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
$\overline{X} \approx 79.64$	$\overline{X} = 79.62$	$\overline{X} = 78.78$	$\overline{X} = 80.52$

$(79.62 - 79.64)^2 \approx 0.0$
$(78.78 - 79.64)^2 \approx 0.74$
$(80.52 - 79.64)^2 \approx 0.77$

S_B^2 = 5x(column sum) 5(1.51) = 7.55

3.

$S_W^2 = S_T^2 - S_B^2 = 65.75 - 7.55 = 58.2$

4.

$$F = \frac{S_B^2 c(r-1)}{S_W^2(c-1)} = \frac{(7.55)3(5-1)}{(58.2)(3-1)}. \approx 0.78$$

5.

$d_2 = c - 1 = 3 - 1 = 2$, degrees of freedom

$d_1 = c(r - 1) = 3(5 - 1) = 12$, degrees of freedom

From Table E, for $\alpha = 0.05$, we find $F_{0.05} = 3.89$.

Since $F = 0.78 < 3.89$, we conclude there is no significant difference in grades between class sections.

6.

Since we have computed the total variation in problem 1 for 1 factor analysis, we can use the same number for 2 factor analysis: $S_T^2 \approx 65.75$.

7.

Latin Section/Average grade	**Section 1**	**Section 2**	**Section 3**	**Row Average**
1991	78.20	76.50	79.20	$\overline{X} \approx 77.97$
1992	79.10	77.90	79.87	$\overline{X} \approx 78.96$
1993	79.60	77.60	78.66	$\overline{X} \approx 78.62$
1994	80.10	81.78	79.10	$\overline{X} \approx 80.33$
1995	81.10	80.10	85.78	$\overline{X} \approx 82.33$
$\overline{X} \approx 79.64$	$\overline{X} \approx 79.62$	$\overline{X} \approx 78.78$	$\overline{X} \approx 80.52$	

(Row average - 79.64)²
$(77.97 - 79.64)^2 \approx 2.79$
$(78.96 - 79.64)^2 \approx 0.46$
$(78.62 - 79.64)^2 \approx 1.04$
$(80.33 - 79.64)^2 \approx 0.48$
$(82.33 - 79.64)^2 \approx 7.24$
$S_R^2 = \text{cSum} = 3(12.01) = 36.03$

8.

The variation between columns is computed in problem 2: $S_C^2 = 7.55$.

9.

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

$$S_E^2 = S_T^2 - S_R^2 - S_C^2 = 65.75 - 36.03 - 7.55 \approx 22.17 .$$

10.

$$F = \frac{(r - 1)S_C^2}{S_E^2} = \frac{(5-1)7.55}{22.17} \approx 1.36$$

$d_2 = c - 1 = 3 - 1 = 2$, degrees of freedom.

$d_1 = (r - 1)(c - 1) = (5 - 1)(3 - 1) = 8$, degrees of freedom.

From Table E, we find $F_{0.05} = 4.46$.

Since $F = 1.36 < 4.46$, there is no significant difference in grades between class sections.

11.

$$F = \frac{(c - 1)S_R^2}{S_E^2} = \frac{(3-1)36.03}{22.17} \approx 3.25$$

$d_2 = r - 1 = 5 - 1 = 4$, degrees of freedom

$d_1 = (r - 1)(c - 1) = (5 - 1)(3 - 1) = 8$, degrees of freedom

From Table E, we find $F_{0.05} = 3.84$.

Since $F = 3.25 < 3.84$, there is no significant difference in grades between the five years. .