MS Module 14: Two-factor ANOVA, interaction effects – practice problems

(The attached PDF file has better formatting.)

Exercise 14.1: Two-factor ANOVA, additive model

	Column 1	Column 2
Row 1	12; 14	16;18
Row 2	20; 22	24; 26

Each cell of the table shows the two observations in that group.

We test whether Row 1 differs from Row 2, whether Column 1 differs from Column 2, and whether interaction effects are significant. The ANOVA table calls the rows the A dimension and the columns the B dimension, following the usage in the textbook.

A. What is the square of the sum of all the observations, or $x_{...2}$?
B. What is the sum of the squares of all the observations, or $\Sigma_i\, \Sigma_j\, \Sigma_k\, x_{ijk2}$?
C. What is the sum of the squares of the totals in each cell, or $\Sigma_i\, \Sigma_j\, x_{ij}^2$?
D. What is the sum of the squares of the row totals, or $\Sigma_j\, x_{i..2}$
E. What is the sum of the squares of the column totals, or $\Sigma_j\, x_{.j2}$
F. What is SST, the total sum of squared deviations?
G. What is SSA, the sum of squared deviations for the i dimension?
H. What is SSB, the sum of squared deviations for the j dimension?
I.   What is SSE, the error sum of squared deviations?
J. What is SSAB, the sum of squared deviations for the interaction?
K. What are the degrees of freedom for the rows (SSA)?
L. What are the degrees of freedom for the columns (SSB)?
M. What are the degrees of freedom for the interaction effects (SSAB)?
N. What are the degrees of freedom for the total sum of squares (SST)?
O. What are the degrees of freedom for the error sum of squares (SSE)?
P. What is MSA, the mean squared deviation for the rows?
Q. What is MSB, the mean squared deviation for the columns?
R. What is MSAB, the mean squared deviation for the interaction?
S. What is MSE, the mean squared error?
T. What is f_A, the f value for testing significance of the row differences?
U. What is f_B, the f value for testing significance of the column differences?
V. What is f_{AB}, the f value for testing significance of the interaction effect?

Part A: The sum of all the observations is

$$12 + 14 + 16 + 18 + 20 + 22 + 24 + 26 = 152$$

The square of this sum is $152^2 = 23{,}104$

This squared sum, which does not differentiate by row or column, is used for the total sum of squares SST.

Part B: The sum of the squares of all the observations is

$$12^2 + 14^2 + 16^2 + 18^2 + 20^2 + 22^2 + 24^2 + 26^2 = 3{,}056$$

Question: What does the sum of squares of each observation tell us? If the two observations in the first cell have an average of 13, does it make a difference if the observations are 12 and 14 vs 2 and 24?

Answer: The significance of the ANOVA depends on the variance of the random error term.

- If the two observations in a cell are similar, such as 12 and 14, 16 and 18, 20 and 22, or 24 and 26, the remaining variance – after adjusting for the differences by rows, columns, and interaction effects – is small (as it is in this practice problem).
- If the two observations in a cell are dis-similar, such as 2 and 24, the remaining variance – after adjusting for the differences by rows, columns, and interaction effects – is large.

This practice problem is heuristic so that you can follow the logic of the two way analysis of variance.

- The observations in each cell differ by 2.
- The columns in each row differ by 4.
- The rows in each column differ by 8.
- The model is additive (see below).

The mean squares differ accordingly:

- The MSE (mean squared error) is 8.
- The MSB (columns) is $8 \times 2^2 = 32$.
- The MSA (rows) is $32 \times 2^2 = 128$.
- The MSAB (interactions) is 0, since the model is additive.

We solved this problem intuitively. The rest of this solution solves the two-way ANOVA by the computation formulas. The next problem changes the input figures so you see the effects on the mean squares.

Part C: The sum of the squares of the totals in each cell, or $\Sigma_i \Sigma_j x_{ij.2}$, is

$$26^2 + 34^2 + 42^2 + 50^2 = 6{,}096$$

This sum of squares does not differentiate by row or column. It is used for the error sum of squares.

Part D: The row totals are 26 + 34 = 60 for Row 1 and 42 + 50 = 92 for Row 2. The sum of squares is

$$60^2 + 92^2 = 12{,}064$$

This sum of squares differentiates by row but not by column, so it is used for SSA.

	Column 1	Column 2	Total	Squared
Row 1	26	34	60	3,600
Row 2	42	50	92	8,464
Total	68	84	152	12,064
Squared	4,624	7,056	11,680	

Part E: The column totals are 26 + 42 = 68 for Column 1 and 34 + 50 = 84 for Column 2. The sum of squares is

$$68^2 + 84^2 = 11{,}680$$

This sum of squares differentiates by column but not by row, so it is used for SSB.

Part F: SST = $\Sigma_i \Sigma_j \Sigma_k x_{ijk2} - x_{...2} / N$ = 3,056 - 23,104 / 8 = 168.00

Question: What is the intuition for this formula?

Answer: The textbook gives a proof for this formula by simplifying the definition of the total sum of squares. The intuition for the formula may be grasped by starting with no random error term (all observations are the same) and then adding random errors.

If all the N observations are the same value Z, the square of the sum is $(N \times Z)^2$ and the sum of the N squares is $N \times Z^2$, so the formula gives zero. If the observations differ from each other, the sum of squares increases.

Illustration: Keep the total $N \times Z$ but increase half the observations by a constant *k* and decrease half the observations by a constant *k*. The sum of the N squares is $\frac{1}{2} N \times (Z+k)^2 + \frac{1}{2} N \times (Z-k)^2 = N \times Z^2 + N \times k^2$.

Part G: SSA =
    the sum of squares of the row totals ÷ (the number of columns × the number of observations per cell)
–   the square of the sum of all the observations ÷ the number of observations =

$$\frac{1}{4} \times 12{,}064 - 23{,}104 / 8 = 128.00$$

Question: What is the intuition for this formula?

Answer: The formula is the same as the formula for the total sum of squares except that we use only the totals of the observations for each row.

Each observation for the row totals is the sum of four individual observations: two columns × two observations in each cell. The sum of squares of row totals looks at the variability between rows, ignoring any variability between columns or between observations in a cell.

Part H: SSB =

    the sum of squares of the columns totals ÷ (the number of rows × the number of observations per cell)
–   the square of the sum of all the observations ÷ the number of observations =

$$\frac{1}{4} \times 11{,}680 - 23{,}104 / 8 = 32.00$$

Part I: SSE, the error sum of squares, $= \sum_i \sum_j \sum_k x_{ijk2} - \sum_i \sum_j x_{ij}^2 / K = 3{,}056 - 6{,}096 / 2 = 8.00$

Question: What is the intuition for the value of 8?

Answer: We derive the SSE by first principles:

- Each observation differs from the mean in the cell by 1.
- The square of this deviation is $1^2 = 1$.
- The sum of the 8 deviations is $8 \times 1 = 8$.

Part J: SSAB, the sum of squared deviations for the interaction, is derived by subtraction.

$$SST = SSA + SSB + SSAB + SSE \Rightarrow$$
$$SSAB = SST - (SSA + SSB + SSE) =$$

$$168.00 - 128.00 - 32.00 - 8.00 = 0.00$$

Question: Why is the sum of squared deviations for the interaction equal to zero? All the observations differ, and all the other sums of squared deviations are positive.

Answer: The ANOVA assumes an additive model: the expected value of a cell =

- the contribution for the row α_i
- the contribution for the column β_j

- the contribution for the interaction γ_{ij}

The row and column contributions are primary; the interaction contribution is secondary. This practice problem is constructed so that the row and column contributions fully account for the average value in each cell. The ANOVA tests whether the interaction effect is significant. For this problem, the interaction effect is zero.

Question: Intuitively, what does it mean that the interaction effect is zero?

Answer: For an additive model, if the interaction effect is zero, the difference between any two rows does not differ by column, and the difference between any two columns does not differ by row.

For this practice problem, column 2 – column 1 =

- Row 1: 34 – 26 = 8
- Row 2: 50 – 42 = 8

The difference between the columns does not differ by row.

Row 2 – row 1 =

- Column 1: 42 – 26 = 16
- Column 2: 50 – 34 = 16

The difference between the rows does not differ by column.

This posting starts with a simplified additive model: no interaction effect. The computation formulas give SST, SSA, SSB, and SSE. We derive SSAB = 0 by subtraction. We then slightly change one of the observations. SST, SSA, SSB, and SSE change slightly, and the interaction effect is positive.

Question: What causes the relative size of SSA, SSB, and SSE?

Answer: In this practice problem, the two observations in any cell differ by 2, so the average deviation is 1.

- The error sum of squares SSE = $8 \times 1^2 = 8$.
- The degrees of freedom for SSE is 4, so the mean squared error = 8 / 4 = 2.

Intuition for SSA and MSA:

- The sums of the two rows differ by 32.
- Each row has two cells (with two observations per cell).
- The means of the cells in the two rows differ by 16, and the average deviation is 8.
- The row sum of squares SSA = $2 \times 8^2 = 128$.
- The degrees of freedom for SSA is 1, so the mean squared error = 128 / 1 = 128.

Intuition for SSB and MSB:

- The sums of the two columns differ by 16.
- Each column has two cells (with two observations per cell).
- The means of the cells in the two columns differ by 8, and the average deviation is 4.
- The column sum of squares SSB = $2 \times 4^2 = 32$.
- The degrees of freedom for SSA is 1, so the mean squared error = 32 / 1 = 32.

Part K: The degrees of freedom for the rows = the number of rows – 1 = 2 – 1 = 1.

Part L: The degrees of freedom for the columns = the number of columns – 1 = 2 – 1 = 1.

Part M: The degrees of freedom for the interaction effects = (rows − 1) × (columns − 1) = 1 × 1 = 1.

Part N: The degrees of freedom for the total sum of squares = the number of observations − 1) = 8 − 1 = 7.

Part O: The degrees of freedom for the total sum of squares = the sum of the degrees of freedom for SSA, SSB, SSAB, and SSE ⇒ the degrees of freedom for SSE = 7 − 1 − 1 − 1 = 4.

Part P: MSA, the mean squared deviation for the rows, is SSA / degrees of freedom = 128 / 1 = 128.

Part Q: MSB, the mean squared deviation for the columns, is SSB / degrees of freedom = 32 / 1 = 32.

Part R: MSAB, the mean squared deviation for the interaction, is SSAB / degrees of freedom = 0 / 1 = 0.

Part S: MSE, the mean squared error, is SSE / degrees of freedom = 8 / 4 = 2.

Part T: The f_A (*f* value for testing significance of the row differences) is MSA / MSE = 128 / 2 = 64. The *p* value is $F_{64,1,4}$ = 0.001324.

Part U: The f_B (*f* value for testing significance of the column differences) is MSB / MSE = 32 / 2 = 16. The *p* value is $F_{16,1,4}$ = 0.01613.

Part V: The f_{AB} (*f* value for testing significance of the interaction) is MSAB / MSE = 0 / 2 = 0. The *p* value is 0.

(The next exercise changes one value – the second observation in row 2 column 2 is 28 instead of 26. The observations are no longer exactly additive, so the results change and the interaction term is no longer zero.)

Exercise 14.2: Interaction effects

A classification table has two rows, two columns, and two observations in each cell:

	Column 1	Column 2
Row 1	12; 14	16;18
Row 2	20; 22	24; **28**

Each cell of the table shows the two observations in that group.

We test whether Row 1 differs from Row 2, whether Column 1 differs from Column 2, and whether interaction effects are significant. The ANOVA table calls the rows the A dimension and the columns the B dimension, following the usage in the textbook.

A. What is the square of the sum of all the observations, or $x_{..2}$?
B. What is the sum of the squares of all the observations, or $\Sigma_i \Sigma_j \Sigma_k x_{ijk2}$?
C. What is the sum of the totals in each cell, or $\Sigma_i \Sigma_j x_{ij}^2$?
D. What is the sum of the squares of the row totals, or $\Sigma_j x_{i.2}$
E. What is the sum of the squares of the column totals, or $\Sigma_j x_{.j.2}$
F. What is SST, the total sum of squared deviations?
G. What is SSA, the sum of squared deviations for the i dimension?
H. What is SSB, the sum of squared deviations for the j dimension?
I. What is SSE, the error sum of squared deviations?
J. What is SSAB, the sum of squared deviations for the interaction?
K. What are the degrees of freedom for the rows (SSA)?
L. What are the degrees of freedom for the columns (SSB)?
M. What are the degrees of freedom for the interaction effects (SSAB)?
N. What are the degrees of freedom for the total sum of squares (SST)?
O. What are the degrees of freedom for the error sum of squares (SSE)?
P. What is MSA, the mean squared deviation for the rows?
Q. What is MSB, the mean squared deviation for the columns?
R. What is MSAB, the mean squared deviation for the interaction?
S. What is MSE, the mean squared error?
T. What is f_A, the f value for testing significance of the row differences?
U. What is f_B, the f value for testing significance of the column differences?
V. What is f_{AB}, the f value for testing significance of the interaction effect?

Part A: The sum of all the observations is

$$12 + 14 + 16 + 18 + 20 + 22 + 24 + 28 = 154$$

The square of this sum is $154^2 = 23{,}716$.

This squared sum, which does not differentiate by row or column, is used for the total sum of squares SST.

Part B: The sum of the squares of all the observations is

$12^2 + 14^2 + 16^2 + 18^2 + 20^2 + 22^2 + 24^2 + 28^2 = 3{,}164$.

Part C: The sum of the squares of the totals in each cell, or $\Sigma_i \Sigma_j x_{ij.2}$, is

$$26^2 + 34^2 + 42^2 + 52^2 = 6,300.$$

This sum of squares does not differentiate by row or column. It is used for the error sum of squares.

Part D: The row totals are 26 + 34 = 60 for Row 1 and 42 + 52 = 94 for Row 2. The sum of squares is

$$60^2 + 94^2 = 12,436.$$

This sum of squares differentiates by row but not by column, so it is used for SSA.

	Column 1	Column 2	Total	Squared
Row 1	26	34	60	3,600
Row 2	42	52	94	8,836
Total	68	86	154	12,436
Squared	4,624	7,396	12,020	

Part E: The column totals are 26 + 42 = 68 for Column 1 and 34 + 52 = 86 for Column 2. The sum of squares is

$$68^2 + 86^2 = 12,020.$$

This sum of squares differentiates by column but not by row, so it is used for SSB.

Part F: SST = $\Sigma_i \Sigma_j \Sigma_k x_{ijk2} - x_{...2} / N$ = 3,056 - 23,104 / 8 = 168.00

Part G: SSA =
    the sum of squares of the row totals ÷ (the number of columns × the number of observations per cell)
–   the square of the sum of all the observations ÷ the number of observations =

$$\frac{1}{4} \times 12,436 - 23,716 / 8 = 144.50$$

Part H: SSB =

    the sum of squares of the columns totals ÷ (the number of rows × the number of observations per cell)
–   the square of the sum of all the observations ÷ the number of observations =

$$\frac{1}{4} \times 12,020 - 23,716 / 8 = 40.50$$

Part I: SSE, the error sum of squares, = $\Sigma_i \Sigma_j \Sigma_k x_{ijk2} - \Sigma_i \Sigma_j x_{ij}^2 / K$ = 3,164 – 6,300 / 2 = 14.00

Part J: SSAB, the sum of squared deviations for the interaction, is derived by subtraction.

$$SST = SSA + SSB + SSAB + SSE \Rightarrow$$
$$SSAB = SST - (SSA + SSB + SSE) =$$

$$199.50 - 144.50 - 40.50 - 14.00 = 0.50$$

Part K: The degrees of freedom for the rows = the number of rows – 1 = 2 – 1 = 1.

Part L: The degrees of freedom for the columns = the number of columns – 1 = 2 – 1 = 1.

Part M: The degrees of freedom for the interaction effects = (rows – 1) × (columns – 1) = 1 × 1 = 1.

Part N: The degrees of freedom for the total sum of squares = the number of observations – 1) = 8 – 1 = 7.

Part O: The degrees of freedom for the total sum of squares = the sum of the degrees of freedom for SSA, SSB, SSAB, and SSE ⟹ the degrees of freedom for SSE = 7 – 1 – 1 – 1 = 4.

Part P: MSA, the mean squared deviation for the rows, is SSA / degrees of freedom = 144.50 / 1 = 144.50

Part Q: MSB, the mean squared deviation for the columns, is SSB / degrees of freedom = 40.50 / 1 = 40.50

Part R: MSAB, the mean squared deviation for the interaction, is SSAB / degrees of freedom = 0.5 / 1 = 0.5

Part S: MSE, the mean squared error, is SSE / degrees of freedom = 14 / 4 = 3.50

Part T: The f_A (f value for testing significance of the row differences) is MSA / MSE = 144.50 / 3.50 = 41.28571

The p value is $F_{41.28571,1,4}$ = 0.003016.

Part U: The f_B (f value for testing significance of column differences) is MSB / MSE = 40.50 / 3.50 = 11.57143

The p value is $F_{11.57143,1,4}$ = 0.027235.

Part V: The f_{AB} (f value for testing significance of the interaction) is MSAB / MSE = 0.5 / 3.5 = 0.142857

The p value is $F_{0.142857,1,4}$ = 0.724659.