Exercise 07 – December 11, 2024

1. A researcher wants to compare the growth of plants under three types of fertilizers (A, B, and C). The heights of the plants after 30 days (in cm) are:

Fertilizer A	Fertilizer B	Fertilizer C
15	20	25
16	22	27
14	19	26
15	21	28
17	20	24

Does the type of fertilizer (A, B, or C) significantly affect plant growth (with $\alpha = 0.05$)?
Perform a one-way ANOVA to determine if fertilizer type affects plant growth.
Create a null hypothesis and alternative hypothesis first.

Solution:

State the Hypotheses

Null Hypothesis (H_0):

The mean plant heights are the same for all three fertilizers:

$\mu_A = \mu_B = \mu_C$

Alternative Hypothesis (H_1):

At least one fertilizer produces a different mean plant height.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Fertilizer A	5	77	15,4	1,3
Fertilizer B	5	102	20,4	1,3
Fertilizer C	5	130	26	2,5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	281,2	2	140,6	82,705882	9,5762E-08	3,885293835
Within Groups	20,4	12	1,7			
Total	301,6	14				

Group Means and Overall Mean

Calculate the grand mean ($\bar{X}$)

$$\bar{X} = \frac{15+16+14+15+17+20+22+19+21+20+25+27+26+28+24}{15} = 20.6$$

Calculate the group means

$\bar{X}_A = \frac{15+16+14+15+17}{5} = 15.4$

$\bar{X}_B = \frac{20+22+19+21+20}{5} = 20.4$

$\bar{X}_C = \frac{25+27+26+28+24}{5} = 26.0$

Sum of Squares

Total Sum of Squares:

$SS_{total} = (15-20.6)^2 + (16-20.6)^2 + \cdots + (24-20.6)^2 = 89.2$

Between-Groups Sum of Squares:

$SS_{between} = 5 \times ((15.4-20.6)^2 + (20.4-20.6)^2 + (26.0-20.6)^2) = 71.6$

Within-Groups Sum of Squares:

$SS_{within} = SS_{total} + SS_{between} = 89.2 - 71.6 = 17.6$

Calculate Mean Squares and F-Statistic

Degree of Freedom (df):

$df_{between} = k-1 = 3-1 = 2$

$df_{within} = N-k = 15-3 = 12$

Mean Squares (MS):

$MS_{between} = SS_{between} / df_{between} = 71.6 / 2 = 35.8$

$MS_{within} = SS_{within} / df_{within} = 17.6 / 12 = 1.47$

Calculate FFF-Statistic:

$F = \frac{\text{MSbetween}}{\text{MSwithin}} = \frac{35.8}{1.47} \approx 24.35$

Decision

Critical F-Value:

From an F-distribution table with $df_{between} = 2$ and $df_{within}=12$ at $\alpha = 0.05$, the critical value is $F_{critical} = 3.89$

Compare F:

Since $F = 24.35 > 3.89$, reject the null hypothesis.

OR

Calculate the p-Value

The p-value is the probability of observing an F-value as extreme as the calculated value ($F = 24.35$) under the null hypothesis.

Using $df_{between} = 2$ and $df_{within} = 12$, the p-value can be found using an F-distribution table or statistical software.

For F = 24.35: Using statistical software or a table, we find that p-value <0.001

Decision Rule

Compare the p-value to α=0.05\alpha = 0.05α=0.05:

- If p ≤ α, reject the null hypothesis (H0H_0H0).
- If p > α, fail to reject the null hypothesis.

In this case, p < 0.001 <0.05, so we **reject the null hypothesis**.

Conclusion

The p-value is extremely small (p < 0.001), which indicates very strong evidence against the null hypothesis. Therefore, we conclude that the type of fertilizer has a significant effect on plant growth.

OR

The F-statistic (F=24.35) is significant at α = 0.05. This indicates that the type of fertilizer has a significant effect on plant growth. At least one fertilizer produces a different mean plant height.

2. A researcher wants to determine if there is an association between **plant type** and **fertilizer preference**. The researcher surveys 90 plants and records the following data:

Fertilizer	Plant Type A	Plant Type B	Plant Type C	Total
Fertilizer X	10	20	10	**40**
Fertilizer Y	15	10	5	**30**
Fertilizer Z	5	5	10	**20**
Total	**30**	**35**	**25**	**90**

Conduct a Chi-Square test of Independence whether plant type and fertilizer preference are independent at α = 0.05.

Solution:

State the Hypotheses

Null Hypothesis (H_0):

Plant type and fertilizer preference are independent.

Alternative Hypothesis (H_1):

Plant type and fertilizer preference are not independent (there is an association)

Calculate the Expected Frequencies

The formula for the expected frequency for a cell is:

$$E_{ij} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

For each cell:

Fertilizer X, Plant Type A:

$$E_{11} = \frac{40 \times 30}{90} = \frac{1200}{90} = 13.33$$

Fertilizer X, Plant Type B:

$$E_{12} = \frac{40 \times 35}{90} = \frac{1400}{90} = 15.56$$

Fertilizer X, Plant Type C:

$$E_{13} = \frac{40 \times 25}{90} = \frac{1000}{90} = 11.11$$

Repeat for all cells to construct the Expected Frequency Table:

Fertilizer	Plant Type A (E)	Plant Type B (E)	Plant Type C (E)	Total
Fertilizer X	13.33	15.56	11.11	**40**
Fertilizer Y	10	11.67	8.33	**30**
Fertilizer Z	6.67	7.78	5.56	**20**
Total	**30**	**35**	**25**	**90**

Compute the Chi-Square Statistic

The formula for the Chi-Square statistic is:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Calculate for each cell:

Fertilizer X, Plant Type A:

$$\frac{(10-13.33)^2}{13.33} = \frac{(-3.33)^2}{13.33} = \frac{11.09}{13.33} = 0.83$$

Fertilizer X, Plant Type B:

$$\frac{(20-15.56)^2}{15.56} = \frac{(4.44)^2}{15.56} = \frac{19.71}{15.56} = 1.27$$

Fertilizer X, Plant Type C:

$$\frac{(10-13.33)^2}{13.33} = \frac{(-3.33)^2}{13.33} = \frac{11.09}{13.33} = 0.11$$

Repeat for all cells. Summing these values gives:

$$\chi^2 = 0.83 + 1.27 + 0.11 + 2.50 + 0.24 + 1.33 + 0.42 + 0.99 + 3.55 = 11.24$$

Degrees of Freedom

The degrees of freedom (df) for a contingency table is:

df = (Number of Rows − 1) × (Number of Columns − 1)

Here:

df = (3 − 1) × (3 − 1) = 2 × 2 = 4

Determine the Critical Value

From the Chi-Square distribution table, for df = 4 and α = 0.05:

$X^2_{critical} = 9.488$

Decision

Compare the test statistics to the critical value:

- If $\chi^2 \leq \chi^2_{critical}$, fail to reject H_0
- If $\chi^2 > \chi^2_{critical}$, reject H_0

Here:

$\chi^2 = 11.24$ and $\chi^2_{critical} = 9.488$

Since 11.24 > 9.488, we **reject the null hypothesis**.

Conclusion

At $\alpha = 0.05$, there is sufficient evidence to conclude that **plant type** and **fertilizer preference** are not independent. There is an association between plant type and fertilizer preference

3. A professor wants to investigate whether the **type of programming language** (Python, Java, C++) and the **study method** (Self-Study, Instructor-Led) affects students' test scores. The professor records the test scores of students after completing a course under each combination of factors.

Language	Self-Study	Instructor-Led
Python	78, 82, 85	90, 88, 92
Java	72, 75, 74	85, 80, 84
C++	65, 68, 70	78, 75, 80

Perform a Two-Way ANOVA to determine if there are significant effects of programming language, study method, or their interaction on test scores.
Create all null hypotheses.
Use $\alpha = 0.05$

Solution:

State Hypotheses:

Main Effect of Programming Language (H_0): Mean test scores are the same across Python, Java, and C++.
Main Effect of Study Method (H_0): Mean test scores are the same for Self-Study and Instructor-Led methods.
Interaction Effect (H_0): There is no interaction between programming language and study method.

Grand Mean ($\bar{X}$) = 78.9444 (average for all 18 values)

Group Means:

Python : 85.8333
Java : 78.3333
C++ : 72.6667
Self-Study : 74.3333
Instructor-Led : 83.5556

Python and Self-Study : 81.6667
Python and Instructor-Led : 90.0
Java and Self-Study : 73.6667
Java and Instructor-Led : 83.0
C++ and Self-Study : 67.6667
C++ and Instructor-Led : 77.6667

Compute Sum of Squares:

Total = $\sum (x_{ij} - \bar{X})^2$

Using $\bar{X}$ = 78.94, calculate for each observation:

Sum of Squares for Factor Programming Language (A):

$SSA = 6 * (85.8333 - 78.9444)^2 + 6 * (78.3333 - 78.94)^2 + 6 * (72.6667 - 78.94)^2$

SSA = 523.4394

Sum of Squares for Factor Study Method (B):

$SSB = 9 * (74.3333 - 78{,}9444)^2 + 9 * (83.5556 - 78{,}9444)^2$

SSB = 382.7222

Sum of Squares Within (Error)

SS Python and Self-Study = $(78 - 81.6667)^2 + (82 - 81.6667)^2 + (85 - 81.6667)^2$ = 24.6664

SS Python and Instructor-Led = 8

SS Java and Self-Study = 4,6667

SS Java and Instructor-Led = 14

SS C++ and Self-Study = 12.6667

SS C++ and Instructor-Led = 12.6667

SSE = 24.6667 + 8 + 4.6667 + 14 + 12.6667 + 12.6667 = 76.6664

Total Sum of Squares

$SSTotal = (78 - 78.9444)^2 + (82 - 78.9444)^2 + \ldots + (80 - 78.9444)^2$

SSTotal = 984.9444

SSInteraction = SS Total – SSA – SSB – SSE

SSInteraction = 984.9444 - 523.4394 - 382.7222 - 76.6664 = 2.1164

Degrees of Freedom:

$df_A = 2$, $df_B = 1$, $df_{interaction} = 2$, $df_{within} = 12$, $df_{Total} = 17$

Mean Squares and FFF-Statistics:

MS_A = SS/df = 523.4394/2 = 261.7197

MS_B = 191.3611

$MS_{A \times B}$ = 1.0852

MS_E = 38.3332

$F_A = MS_A / MS_E$ = 261.7197 / 38.3332 = 40.9652

F_B = 59.9045

$F_{A \times B} = 0.1656$

Decision:

p-value Programming Language for

F = 40.965, df = (2,12) at α = 0.05 is 0.00000435

p-value Study Method for

F = 59.9045, df = (1,12) at α = 0.05 is 0.00000527

p-value Interaction for

F = 0.1656, df = (2,12) at α = 0.05 is 0.84928886

Conclusion:

Significant main effects of programming language on test scores. (p-value < α)

Significant main effects of study method on test scores. p-value < α)

No Significant interaction between language and study methods. p-value > α)

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Sample (study)	523,4444444	2	261,7222222	40,9652174	4,3476E-06
Columns (program)	382,7222222	1	382,7222222	59,9043478	5,26602E-06
Interaction	2,111111111	2	1,055555556	0,16521739	0,849605144
Within	76,66666667	12	6,388888889		
Total	984,9444444	17			