Final Exam Preparation 13-01-2025

Week II : visualization

A. Bar & Pie Chart [Qual.] + Line graph

B. Stem & leaf

- sample :

1998		2000
1 1	4	
	3	7
3 3 2	3	2 3 3
8 8 6 5	2	8 8 9
4 3 3 1 1 1 0	2	0 0 1 1 1 2 2 2 3
9 8 7 7 7 6 6 6 5	1	5 6 8 8 8 8 9 9
3 2 1	1	2 2 4 4 4
7	0	6 9

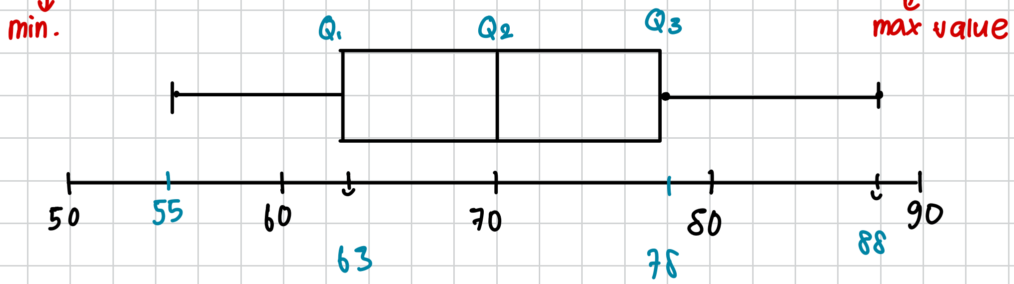
key : 2|3|24
= 1998 : 32
= 2000 : 32
: 34

C. Box Plot :

Q_2

55, 60, 62, 63, 65, 66, 68, 70, 72, 75, 77, 78, 80, 85, 88

min.   max value



→ outliers :

IQR : $Q_3 - Q_1$
: 78 - 63
: 15

→ lower fence : $Q_1 - (1.5 \cdot IQR)$
: 63 - (1.5 × 15)
: 63 - 22.5
: $\underline{\underline{41.5}}$

→

Week III : Permutation & Combination

① Permutation :

> select & arrange ··· how many diff. ways to arrange

$n = 8 \quad r = 4$ (arrange people to)

FORMULA :

$$P(n,r) = \frac{n!}{(n-r)!}$$

$$\rightarrow P(8,4) = \frac{8!}{(8-4)!} = \frac{8!}{4!}$$

$= 1,680$

∴ there are 1,680 ways to arrange (r) out of (n) in a row

② Combination

→ choose, select

$n = 7 \quad r = 4$

FORMULA :

$${}^nC_r = \frac{n!}{(n-r)!\,r!}$$

same but $\frac{x}{x\ r!}$

$$c(7,4) = \frac{7!}{4(7-4)!} = \frac{7!}{4! \times 3!} = \frac{840}{24} = 35$$

OR

$$\frac{7 \times 6 \times 5 \times \cancel{4!}}{\cancel{4!} \times 3!} = \frac{210}{6} = 35$$

(more valid)

∴ there are 35 ways to choose 4

③ red = 10 • random select 5 balls w/t replacement

blue = 15 ↳ prob. of exactly 3 of 5 : red

$N = 25$

$k = 10$ (red)

$n = 5$ (sample)

$x = 3$ (3/5 RED)

$$\rightarrow p : \frac{\text{from } {}_{10}C_3 \times {}_{(N-k)}C_{(5-3)}}{{}_{25}C_5}$$

from all red balls p: 3 red balls

from p: all blue balls

choose 2 blue

all samples

$${}_{10}C_3 = \frac{10!}{7!\,3!} = \frac{10 \times 9 \times 8 \times \cancel{7!}}{\cancel{7!} \times 3!} = \frac{720}{6} = 120$$

(10 - 3)!

$${}_{15}C_2 = \frac{15!}{13! \times 2!} = \frac{15 \times 14 \times \cancel{13!}}{\cancel{13!} \times 2!} = \frac{210}{2} = 105$$

$${}_{25}C_5 = \frac{25!}{20! \times 5!} = \frac{25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot \cancel{20!}}{\cancel{20!} \cdot 5!} = \frac{6\,375\,60\cancel{0}}{12\cancel{0}} = 53,130$$

$$\Longrightarrow \frac{120 \times 105}{53\,130} = 0.237$$

∴ probability of drawing 3 red balls are approx. = 0.2372 or 23.72%

4 s.f. ROUNDED

WEEK 4 :

- % returns
- comparing dist. (visualization)
- Pearson.

① Percentage Returns

year(s) :

1	:	10 %
2	:	15 %
3	:	- 5 %
4	:	8 %
5	:	12 %

a. % → decimal & add ①

1.10
1.15
0.95
1.08
1.12

b. geometric mean formula

$$GM = \sqrt[n]{\underset{\text{multiply}}{\Sigma}\ 1 + x}$$

$$= \sqrt[5]{1.10 \times 1.15 \times 0.95 \times 1.08 \times 1.12}$$

$$= \sqrt[5]{1.422} \approx 1.073$$

c. convert back to %

$$(1.073 - 1) \times 100 = 7.3\ \% \ //$$

∴ the percentage returns from an investment over 5 consecutive years are 7.3%. //

② Box-Plot [comparing distribution]

~ · from 2 different groups ..

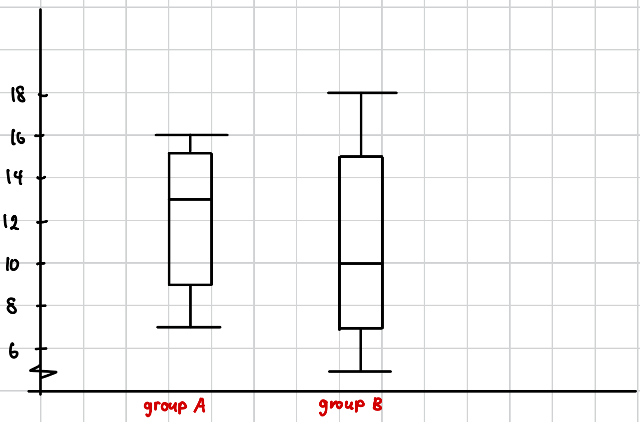
→ compare :

- central tendency
- spread
- potential spread

Group A : 7, 9, 12, 13, 14, 15, 16
Group B : 5, 7, 8, 10, 12, 15, 18

TASKS :

a. calculate 5-number summary (min, Q1, Q2, Q3, max)
b. Draw both box
c. Compare dist. 2 groups based box plot :
   - which group has higher median ?
   - outliers ?



∴ comparing between groups :

→ median : group A median -> group B median

→ outliers : Neither group has extreme outliers based on provided.

Solution :

GROUP A : 7, 9, 12, 13, 14, 15, 16
lower   upper

① · min = 7
  · Q_1 = median of lower half = 9
③ · Q_2 = 13
  · Q_3 = median of upper -"- = 15
  · max = 16

GROUP B : 5, 7, 8, 10, 12, 15, 18

① · min = 5
  · Q_1 = 7
② · Q_2 = 10
  · Q_3 = 15
  · max = 18

③ Card $n = 52$ → flip coin

p: draw king ($4/52$) → ($1/13$)   p: tail ($1/2$)

- two p : independent

  ↳ both happening $P(\text{king \& Tail}) = \frac{1}{13} \times \frac{1}{2} = \frac{1}{26}$ //

④ month sales from 10 - each department.

- Department X Sales : 12, 14, 17, 19, 21, 24, 26, 28, 30, 32
- Department Y Sales : 13, 16, 18, 20, 23, 25, 27, 29, 31, 33

- construct back2back Stem & Leaf :

Dep X	stem	Dep Y
3 6 8	1	2 4 7 9
0 3 5 7 9	2	1 4 6 8
1 3	3	0 2

⑤ P : 3 heads , flip 5x (heads = success)

Solution :

$N = 5$

$x = 3$

$\pi = \frac{1}{2} = 0.5$

Formula :

$$P(x=3) = \frac{N!}{x!\,(N-x)!}\,\pi^x (1-\pi)^{N-x}$$

$$= \frac{5!}{3!\,(5-3)!}\,\pi^x (1-\pi)^{5-3}$$

$$= \frac{5 \times 4 \times \cancel{3!}}{\cancel{3!} \times 2!} = \frac{20}{2} = 10$$

$$\Rightarrow 10\,(0.5)^3 (0.5)^2$$

$$\Rightarrow 10 \left(\tfrac{1}{2}\right)^3 \left(\tfrac{1}{2}\right)^2$$

$$\Rightarrow 10 \times \frac{1}{8} \times \frac{1}{4}$$

$$10 \times \frac{1}{32} = \frac{10}{32} = \frac{5}{16} \text{ // OR } 0.3125 \text{ //}$$

⑥ p success : 80%

$N : 15$

p of ≥ 12 successful free throws.

$P(x \geq 12) = \sum P(x = 12, 13, 14, 15)$

limit : no need normal distribution.

$$P(x=12) = \frac{15!}{(15-3)!\,3!}\,(\pi)^{12}(1-\pi)^3$$

$$= \frac{15 \times 14 \times 13 \times \cancel{12!}}{\cancel{12!} \times 3!}$$

$$= \frac{15 \times 14 \times 13}{3!} \times 0.8^{12} \times 0.2^3 = 0.227$$

$P(x=13) = 0.236$

$P(x=14) = 0.137$

$P(x=15) = 0.035$

$\therefore$ SUM $= 0.635$ //

⑦ Biologist

relationship between : Pearson Correlation Coeff.

X : hour(s) of sunlight

Y : Height (cm)

X	y
2	10
4	15
6	20
8	25
10	30

• Calculate Pearson's Coefficient :

	X	Y	x	y	xy	x^2	y^2
	2	10	-4	-10	40	16	100
	4	15	-2	-5	10	4	25
	6	20	0	0	0	0	0
	8	25	2	5	10	4	25
	10	30	4	10	40	16	100
total	30	100	0	0	100	40	250
Mean	6	20	0	0			

Formula :

$r = (\Sigma xy / \Sigma x^2 \Sigma y^2) \rightarrow 100 / \sqrt{40 + 250}$

$r = 100 / 100 = 1$

Week # 5

① Standard Deviation

: 70, 85, 78, 90, 88

$$\bar{x} = \frac{70+85+78+90+88}{5} = 82.2$$

- deviation from $\bar{x}$ per score
  - $70 - 82.2 = -12.2$
  - $85 - 82.2 = 2.8$
  - $78 - 82.2 = -4.2$
  - $90 - 82.2 = 7.8$
  - $88 - 82.2 = 5.8$

- squared deviations
  - $(-12.2)^2 = 148.84$
  - $(2.8)^2 = 7.84$
  - $(-4.2)^2 = 17.64$
  - $(7.8)^2 = 60.84$
  - $(5.8)^2 = 33.64$

- Variance ($\bar{x}$ of squared deviation)

$$= \frac{148.84 + 7.84 + 17.64 + 60.84 + 33.64}{5} = 53.76$$

- standard deviation $\sqrt{var.}$

$$= \sqrt{53.76} \approx 7.33$$

② 30% prefer coffee → Normal Distribution

select = 100

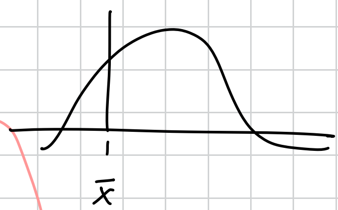
$p < 25$ prefer coffee : use z-table

a. $n = 100 \quad p = 0.30 \quad q = 0.70$

b. check conditions

- $n \cdot p = 100 \times 0.30 = 30$
- $n \cdot (1-p) = 100 \times 0.70 = 70$

: more than 5 : big area



> 5

c. $\bar{x}$ & sd

- $\mu = 100 \cdot 0.30 = 30$
- $\sigma = \sqrt{100 \cdot 0.30 \cdot 0.70} = \sqrt{21} \approx 4.58$

approximation since

d. continuity correction

$P(X < 25) \rightarrow P(X \leq 24.5$

e. Standardize :

$$z = \frac{\overset{\text{approx}}{24.5} - \overset{\mu}{30}}{4.58} = \frac{-5.5}{4.58} \approx -1.20$$

→ find z-score :

: ≈ 0.1151 OR 11.51%

③ $n = 100$ times → Normal Distribution

p success $= 0.4$

$p_{success} \geq 45$?

a. $n \cdot p = 100 \cdot 0.4 = 40$

$n(1-p) = 100 \cdot 0.6 = 60$

b. $\bar{x}$ & sd.

$\mu = 100 \cdot 0.4 = 40$

$\sigma = \sqrt{100 \cdot 0.4 \cdot 0.6} = \sqrt{24} \approx 4.9$

c. apply continuity correction

$P(X \geq 44.5)$

d. z score approx

$$z = \frac{44.5 - 40}{4.9} = \frac{4.5}{4.9} \approx 0.92$$

e. p z-table $= P(z \leq 0.92) \approx 0.8212$

$1 - 0.8212 = 0.1788$ // OR 17.88% //

Week 6 : null hypothesis
: test.

① sig. diff. of $\bar{x}$ from 1000 hours using $\alpha = 0.05$

- null hypo. H_0 = mean lifespan is 1000 hours using $\alpha = 0.05$
- alternative H_1 = mean lifespan is not 1000 hours ($\mu = 1000$)

find mean

Sample μ : $\dfrac{950 + 900 + 970 + 980 + 1020 + 1030 + 990 + 1010 + 1000 + 995}{n = 10} = 990.5$ (Σ values)

sd : $s \approx 25.87$
FORMULA :

A. t-statistic :

$$t = \frac{\overset{\text{sample}}{\bar{y}} - \overset{\text{expected}}{x}}{sd / \sqrt{n}} = \frac{990.5 - 1000}{25.87 / \sqrt{10}} \approx \frac{-9.5}{8.18} \approx -1.16$$ } initial t : BEFORE

B. comparing t-stats w/ critical value

DOF : n - 1 → 10 - 1 = 9

↳ at $\alpha = 0.05$

t : -1.16 — t. dist. table is within range [-2.262, 2.262] OR ±2.262

critical t value > α 0.05 = fail to reject null hypothesis.

dof : 9 — 2.262 (α 0.05), since t = -1.16 is within range = fail to reject null hypothesis.

② weight of n clients

· Before & After (kg) training program.

Client	Before (kg)	After (kg)	Difference (d)
1	85	82	-3
2	78	75	-3
3	90	85	-5
4	76	74	-2
5	88	85	-3
6	81	78	-3
7	79	76	-3
8	92	89	-3

Conduct paired t-test

↳ see significant reduction in weight

$\alpha = 0.05$

H_0 = traing program has no effect on weight ($\mu_d = 0$)

H_1 = training program reduces weight ($\mu_d < 0$) less than sample (critical t)

-3 - -3.125

-3 + 3.125

$(0.125)^2$

A. $\bar{d} = \dfrac{\Sigma d}{n} = \dfrac{-25}{8} = -3.125$

mean

substract $\bar{x}$

d_i : each data point

sum OF ALL $(d_i - \bar{x})^2$

$S_d = \sqrt{\dfrac{\Sigma(d_i - \bar{d})^2}{n-1}} \approx 0.835$

dof = n - 1 = 7 → t statistic : $t = \dfrac{\bar{d}}{s_d / \sqrt{n}} = \dfrac{-3.125}{0.835 / \sqrt{8}} = \dfrac{-3.125}{0.295} \approx -10.59$

variance

Critical t-value

: $\alpha = 0.05$ (one-tailed)

$= 1.895$

: since $-10.59 < 1.895 \rightarrow$ null hypothesis H_0 : rejected

③ test if new diet plan (Group A) significantly improves weight <u>loss</u> compared to standard (Group B)

→ Q : perform independent t-test significantly improve weight loss at level $\alpha = 0.05$

Group	Sample Size (n)	Mean Weight Loss (x)	Standard Deviation (s)
Group A (New)	25	8 kg	2
Group B (Standard)	25	6 kg	2.5

$=$ $= \bar{x}$ $=$

H_0 = μ weight for BOTH groups are equal ($\mu_A = \mu_B$)

H_1 = new diet plan : greater weight loss ($\mu_A > \mu_B$)

one tailed

Calculate t-stats. (Assume greater)

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s^2_A}{n_A} + \frac{s^2_B}{n_B}}}$$

$$\frac{A - B}{\text{variance } (Sd^2) \text{ SUM } A + B}$$

$$= \frac{8 - 6}{\sqrt{\frac{2^2}{25} + \frac{2.5^2}{25}}} = \frac{2}{\sqrt{0.16 + 0.25}}$$

$$= \frac{2}{\sqrt{0.41}} \approx \frac{2}{0.64} \approx 3.13$$

DOF & critical t-value

$dof = n_A + n_B - 2 = 25 + 25 - 2 = 48$

$(n_1 + n_2 - 1 - 1)$

at $\alpha = 0.05$ (one tailed) : closest value 1.697

or

calculator = 1.7 something

since $3.13 > 1.679$, H_0 is rejected.

, new diet plan leads to significantly higher weight loss

Week 7

① one-way ANOVA

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 α = 0.05)?
Perform a one-way ANOVA to determine if fertilizer type affects plant growth.
Create a null hypothesis and alternative hypothesis first.

H_0 : mean plant heights are same for all 3 fertilizers

$\mu_A = \mu_B = \mu_C$

H_1 : at least one fertilizer produce different mean heights

SUMMARY :

Fertilizer	count	mean	variance
A	5	15.4	1.3
B	5	20.4	1.3
C	5	26	2.5

$k = 3$

ANOVA :

Source of variation	SS	dof df	MS	F	P value	Fcrit.
B	89.2	2	140.6	82.7	9.58	3.89
W	71.6	12	1.7			
total	89.2	14				

grand Mean

$\frac{\sum \text{all}^{\text{sum}}}{n \text{ all}} = 20.6$

- Sum of Squares

all values - great mean

$SS_{total} = 5 \times ((15 - 20.6)^2 + (16 - 20.6)^2 + \dots$ pokoknya jadi 89.2.

count

$SS_B = 5 \times ((15.4 - 20.6)^2 + (20.4 - 20.6)^2 + (26.0 - 20.6)^2)$
$= 71.6$

$SS_W = SS\ total - SS\ between = 89.2 - 71.6 = 17.6$

- DOF

DF Between $= k - 1 = 3 - 1 = 2$
DF Within $= N - k = 15 - 3 = 12$

- Mean Squares (MS) :

$MS_B = SS_B / df_B = 71.6 / 2 = 35.8$
$MS_W = SS_W / df_W = 17.6 / 12 = 1.47$

- FFF- Statistic

$F = \frac{MS_B}{MS_W} = \frac{35.8}{1.47} \approx 24.35$

$df1$ = Between
$df2$ = Within
F-critical = 3.89 (3 s.f.)
since $F = 24.35 > 3.89$, H_0 rejected
F statistic is significant at $\alpha = 0.05$.
indicates significant effect on plant growth

OR

p-value of df_B 2 and $df_W = 12$
→ use CALCULATOR
p value < 0.001

Compare
p value to $\alpha = 0.05$
$0.05\alpha = 0.05$
if $p \le \alpha$ = reject null
$p > \alpha$ = fail
in this case $p < 0.001 < 0.05$
REJECT

② Chi-Square Test

association between plant-type & fertilizer preference

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.

H_0 = plant type & fertilizer = Independent

H_1 = plant type & fertilizer preference are not independent (association)

Ⓐ Calculate Expected Frequencies

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

→ E_{11} (row one column 1) $= \frac{40 \cdot 30}{90} = 13.33$

$E_{12} = \frac{40 \times 35}{90} = 15.56$ (2 sig fig)

→ Expected Frequencies Table

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**

Chi-Square Statistic

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

i = row, j = column, O : original, E : expected

- Calculate For Each Cell

FX, type A :

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

original − expected

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

- DOF for contingency table

= (num of rows − 1) × (num of columns − 1)

= (3 − 1) · (3 − 1) = 2 × 2 = 4

- Crit Value

df = 4 α = 0.05 from Chi² Dist. Table

X^2 critical = 9.488

CORRELATED

Decision

If $x^2 \le x^2_{crit}$ = FAIL

if $x^2 > x^2$ = reject H_0

→ 11.24 > 9.488 , null hypothesis reject

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.

Two way ANOVA

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 α = 0.05

null : H_0
H_0 = Mean Test Scores same amongst all Programming Language
H_0 = Mean Test Score Same for all Study method
H_0 = no interaction between programming & Study method.

Grand Mean = 78.9444

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

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

$SSTotal = (78 – 78.9444)^2 + (82 – 78.9444)^2 + . . . + (80 – 78.9444)^2$

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:

$MS_A = SS/df = 523.4394/2 = 261.7197$

$MS_B = 191.3611$

$MS_{A*B} = 1.0852$

$MS_E = 38.3332$

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

$F_B = 59.9045$

$F_{A*B} = 0.1656$

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			

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