EXERCISE 1

1. Create a Stem-and-Leaf Display

Data set:

62, 65, 68, 70, 73, 75, 75, 78, 81, 83, 84, 85, 87, 89, 92, 95, 96, 98, 100

Stem	Leaf
6	2 5 8
7	0 3 5 5 8
8	1 3 4 5 7 9
9	2 5 6 8
10	0

2. Construct a Box Plot

Given the following datasets of students' test scores:

Dataset:

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

Tasks:

a. Determine the five-number summary (minimum, 25th quartile, 50th quartile, 75th quartile, and maximum)
b. Draw the box plot based on the five-number summary with whiskers (use 1.5 * H-spread to identify outliers)
c. Identify any potential outliers (outside value or/and far out value)

2) a) minimum = 55

25^{th} Quartile = $\frac{n+1}{4} = \frac{15+1}{4} = \frac{16}{4} = 4^{th}$ item = $\underline{\underline{63}}$

50^{th} Quartile = $\underline{\underline{70}}$

75^{th} Quartile = $\frac{3(n+1)}{4} = \frac{3(16)}{4} = \frac{48}{4} = 12^{th}$ item = $\underline{\underline{78}}$

maximum = 88

b) H-spread = $Q_3 - Q_1$ = 78 - 63 = 15

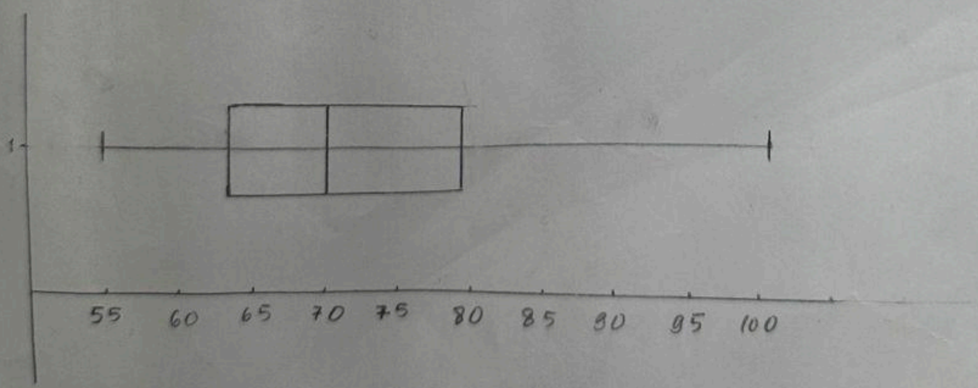
Step = 1.5 x 15 = 22.5

Upper inner fence = Q_3 + 1 step = 78 + 22.5 = 100.5

Lower inner fence = Q_1 - 1 step = 63 - 22.5 = 40.5

Upper adjacent = 88

Inner adjacent = 55



c) No outliers bcs all data is within 40.5 to 100.5.

EXERCISE 2

1. Calculate the trimean for the dataset below:

   10, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 50

1) Trimean $= \dfrac{(Q_1 + 2Q_2 + Q_3)}{4} = \dfrac{18 + 2(30) + 42}{4} = \dfrac{120}{4} = 30$

$Q_1 = \dfrac{16}{4} = 4^{th}$ term $= 18$

$Q_2 = 30$

$Q_3 = \dfrac{3(16)}{4} = 12^{th}$ term $= 42$

Trimean provides a balanced estimate of central tendency that is robust to outliers.

2. Calculate the geometric mean of the growth rates to find the average population growth rate over these 4 years:.
   - Year 1 = +5%
   - Year 2 = +10%
   - Year 3 = -3%
   - Year 4 = +6%

2) $x = 1.05, 1.10, 0.97, 1.06$

$\bar{x}_{geom} = \sqrt[4]{1.05 \times 1.10 \times 0.97 \times 1.06}$

$= 1.044$

$1.044 - 1 = 0.044 = 4.4\%$ per year.

3. Trimmed mean, consider the following dataset of 10 values representing exam scores (10% trim)

   65, 70, 72, 75, 80, 85, 90, 92, 95, 100

10% × 10 = 1, so remove 1 smallest and 1 largest values.

70, 72, 75, 80, 85, 90, 92, 95

10% trimmed mean = $\frac{659}{8}$ = 82.375.

EXERCISE 3

1. You have 8 people, and you arranged 4 of them in a row for a photo, how many different ways can you arrange them?

"different ways" => combination

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

$$8P4 = \frac{8!}{(8-4)!} = \frac{8 \times 7 \times 6 \times 5 \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{\cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}$$

$$= 1680 \text{ ways}$$

2. You have 7 books, and you want to choose 4 to take on a trip. How many different ways can you select books?

2) $$7C4 = \frac{7!}{(7-4)!\,4!} = \frac{7 \cdot 6 \cdot \cancel{4!}}{3!\,\cancel{4!}} = 35 \text{ ways.}$$

3. A bag contains 10 red, 15 blue balls. If you randomly select 5 balls without replacement, what is the probability that exactly 3 of the selected balls are red?

= 2.191

Hypergeometric formula : $p = \frac{{}_{k}C_{x} \; {}_{(N-k)}C_{(n-x)}}{{}_{N}C_{n}} = \frac{{}_{10}C_{3} \; {}_{15}C_{2}}{{}_{25}C_{5}}$

k = total of color we're looking

x = selected no. of color

N = total no of balls

n = total no. of selection.

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

$= \frac{12\,600}{53\,130} = 0.237$

EXERCISE 4

1. Find the percentage returns from an investment over 5 consecutive years were:
    - Year 1: 10%
    - Year 2: 15%
    - Year 3: -5%
    - Year 4: 8%
    - Year 5: 12%

$$\bar{x}_{geom} = \sqrt[5]{1.1 \times 1.15 \times 0.95 \times 1.08 \times 1.12}$$

$$\approx 1.078$$

$$(1.078 - 1) \times 100\% = 7.8\%$$

2. Create a box plot to compare the distribution of data from two different groups, each containing an odd number of data points. Interpret the box plots to compare the central tendency, spread, and potential outliers between the groups.

   You are given the following data sets for two groups:
    - Group A: 7, 9, 12, 13, 14, 15, 16
    - Group B: 5, 7, 8, 10, 12, 15, 18

   Tasks:
    a. Calculate the five-number summary (minimum, 1st quartile Q1, median, 3rd quartile Q3, and maximum) for each group.
    b. Draw the box plots for both groups on the same axis, labelling the minimum, Q1, median, Q3, and maximum values.
    c. Compare the distributions of the two groups based on the box plots:
        - Which group has a higher median?
        - Are there any outliers

2) a) min = 7

$Q_1 = \frac{7+1}{4} = 2^{nd}$ term = 9

$Q_2 = 13$

$Q_3 = \frac{3(8)}{4} = 6^{th}$ term = 15

max = 16

H-spread = $Q_3 - Q_1 = 15 - 9 = 6$

Step = $1.5 \times 6 = 9$

Upper inner = $15 + 9 = 24$

Lower inner = $9 - 9 = 0$.

b) min = 5

$Q_1 = \frac{8}{4} = 2^{nd}$ term = 7

$Q_2 = 10$

$Q_3 = \frac{3(8)}{4} = 6^{th}$ term = 15

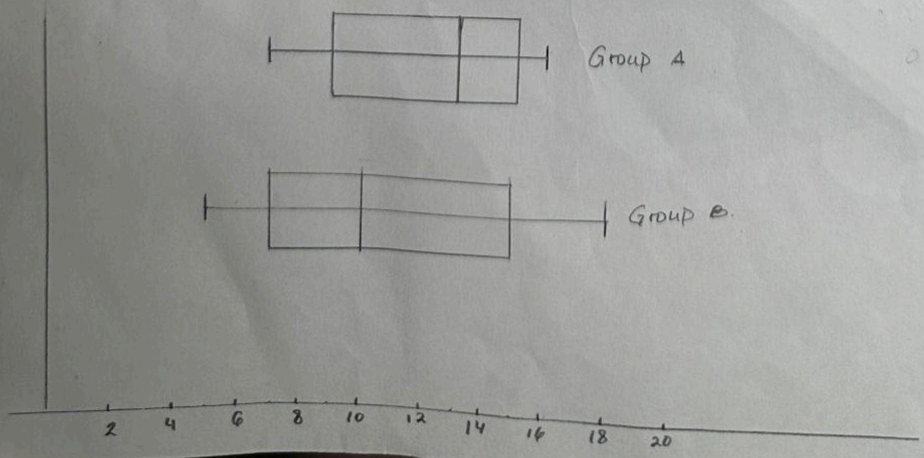
max = 18

H-spred = $15 - 7 = 8$

Step = $1.5 \times 8 = 12$

Upper inner = $15 + 12 = 27$

Lower inner = $7 - 12 = -5$



3. A card is drawn from a standard deck of 52 cards, and then a coin is flipped. What is the probability of drawing a “King” from the deck and flipping a “Tail”?

Independent

$$P(A \text{ and } B) = \frac{\cancel{4}^{\,1}}{\cancel{52}_{\,13}} \times \frac{1}{2} = \frac{1}{26}$$

4. Two departments at a company recorded the number of sales made by their top 10 salespeople in a month. The number of sales made are as follows”
    - 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

    Please, construct a back-to-back stem-and-leaf display for the two datas.

4)

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

5. Calculate the probability of getting exactly 3 heads when flipping a fair coin 5 times (where getting heads is considered a success)

5) Binomial probability (when mentioned success)

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

N = no. of trials

x = no of successes

π = probability of success

$$= \frac{5!}{3!(5-3)!} 0.5^{3}(1-0.5)^{5-3}$$

$$= \frac{5}{16}$$

6. In a basketball game, a player has a free throw success rate of 80%. If the player takes 15 free throws, what is the probability that they make at least 12 successful free throws?

6) $P(x=12) = \frac{N!}{x!(N-x)!} \pi^{x}(1-\pi)^{N-x}$

$$= \frac{15!}{12!(3!)} 0.8^{12}(0.2)^{3}$$

$$= 0.227$$

$P(x=13) = \frac{15!}{13!\ 2!} (0.8)^{13}(0.2)^{2}$

$$= 0.236$$

$P(x=14) = \frac{15!}{14!\ 1!} (0.8)^{14}(0.2)^{1}$

$$= 0.137$$

$P(x=15) = \frac{\cancel{15!}}{\cancel{15!}} 0.8^{15}(0.2)^{0}$

$$= 0.035$$

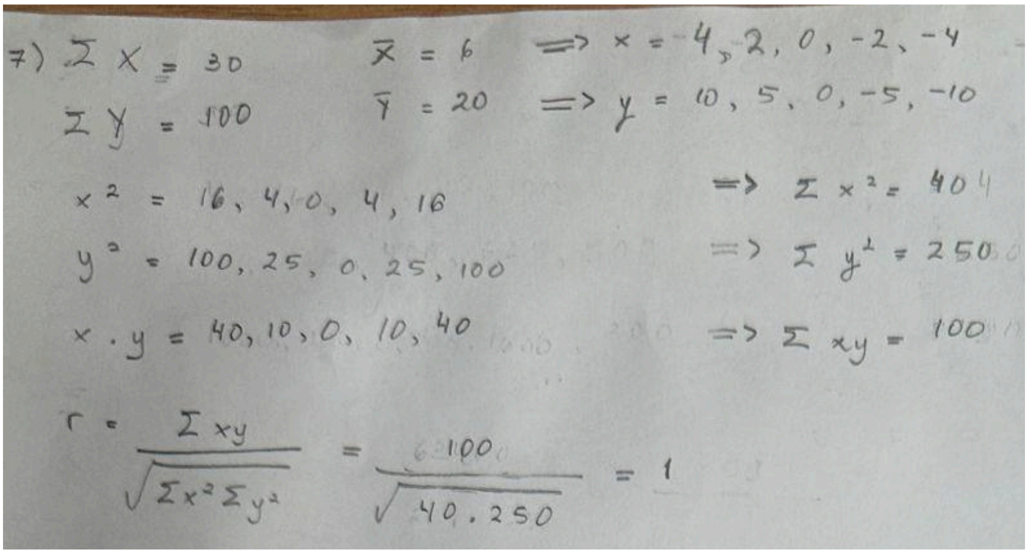
$P(x \geq 12) = P(x=12) + P(x=13) + P(x=14) + P(x=15)$

$$= 0.635$$

7. A biologist studies the relationship between the number of hours of sunlight a plant receives and its height. THe following data shows the hours of sunlight and the corresponding heights of 5 plants:

Calculate the Pearson correlation coefficient.

Hours of Sunlight (X)	Height (cm) (y)
2	10
4	15
6	20
8	25
10	30



7) $\Sigma x = 30$ $\quad \bar{x} = 6 \Rightarrow x = 4, 2, 0, -2, -4$

$\Sigma y = 100$ $\quad \bar{y} = 20 \Rightarrow y = 10, 5, 0, -5, -10$

$x^2 = 16, 4, 0, 4, 16 \Rightarrow \Sigma x^2 = 40$

$y^2 = 100, 25, 0, 25, 100 \Rightarrow \Sigma y^2 = 250$

$x \cdot y = 40, 10, 0, 10, 40 \Rightarrow \Sigma xy = 100$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{100}{\sqrt{40 \cdot 250}} = 1$$

EXERCISE 5

1. Find the standard deviation from these data
   Scores: 70, 85, 78, 90, 88

1) mean = 82.2

$(x_i - \text{mean})^2 \Rightarrow (70 - 82.2)^2 = 148.84$

$(85 - 82.2)^2 = 7.84$

$(78 - 82.2)^2 = 17.64$

$(90 - 82.2)^2 = 60.84$

$(88 - 82.2)^2 = 33.64$ +

268.8

$SD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$ => use n for regular SD (population)
use n − 1 for t-test. (sample)

$= \sqrt{\frac{1}{5}(268.8)}$

$= 7.33$

2. Suppose a survey indicates that 30% of people prefer coffee over tea. If you randomly select 100 people. What is the probability that fewer than 25 people prefer coffee? Use the z table!

2) $n \cdot p \geq 5$
$n(1-p) \geq 5$ } Conditions.

$np = 100 \times 0.30 = 30$

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

mean $= n \times p = 100 \times 0.30 = 30$

SD $= \sqrt{np(1-p)}$

$= \sqrt{100(0.30)(0.70)}$

$= 4.583$

n = no of trials
X = no of successes
p = probability of success.

continuity correction

$P(x < 25) \rightarrow 24 \rightarrow 24.5 = P(x \leq 24.5)$

z score $= \frac{x - \text{mean}}{SD} = \frac{24.5 - 30}{4.583} = -1.200$

z-table gives $P(z \leq -1.20) = 0.1151$

The probability that fewer than 25 people coffee is 0.1151 or 11.51 %

3. You are conducting an experiment with 100 trials (n=100) and the probability of success in each trial is p = 0.4 find the probability that at least 45 success will occur.

3) Check condition

$np = 100 \times 0.4 = 40$
$n(1-p) = 100 \times 0.6 = 60$
} greater than 5.

mean $= 40$

$\sigma = \sqrt{100 \times 0.4 \times 0.6} = 4.899$

$P(x \geq 45) \Rightarrow P(x \geq 44.5)$

$z = \dfrac{44.5 - 40}{4.899} = 0.919$

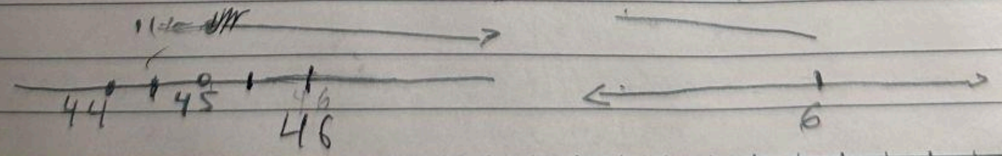
$P(z \leq 0.919) = 0.81594$

Since we're looking for $P(x \geq 45)$, which corresponds to $P(z \geq 0.919)$:

$P(z \geq 0.919) = 1 - P(z \leq 0.919)$
$= 1 - 0.81594$
$= 0.1788 = 17.88\%$

Note :
If doesn't meet checking, use Binomial distribution.

$P(x \geq 45)$ $\qquad$ $P(x < 6)$
$\qquad\qquad\qquad$ $P(x \leq 5.5)$



Continuity correction

EXERCISE 6

1. A company claims their light bulbs last 1000 hours on average. A sample of 10 bulbs yields the following lifespans (in hours):
   950, 960, 970, 980, 1020, 1030, 990, 1010, 1000, 995
   Test whether the mean lifespan differs significantly from 1000 hours using α = 0.05

1) t-table

H_0 : Mean lifespan doesn't differ significantly from 1000 hours

H_1 : Mean lifespan differ significantly from 100 hours

$$\text{mean} = \frac{9905}{10} = 990.5$$

$$\sigma = 25.868$$

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{990.5 - 1000}{\frac{25.868}{\sqrt{10}}}$$

$$= -1.161$$

$$df = n - 1 = 10 - 1 = 9$$

At $\alpha = 0.05$ (two tailed), t-crit is ± 2.262

Since -1.161 is within -2.262 and 2.262, we fail to reject H_0.

2. A fitness coach measures the weight of 8 clients before and after a 6-week 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 a paired t-test to determine if the training program significantly reduced weight. Use α = 0.05

2) H_0 : The 2 population means are equal

H_1: Before is greater than after weight (one tailed)

(left tailed).

mean diff = -3.125

SD diff = 0.835 (sample)

n = 8

$$t = \frac{\text{mean diff}}{\frac{\text{SD diff}}{\sqrt{n}}} = \frac{-3.125}{\frac{0.835}{\sqrt{8}}} = -10.591$$

df = n - 1 = 7

t-value = 1.895

Since -10.591 < 1.895, H_0 is rejected.

3. A nutritionist wants to test if a new diet plan (Group A) significantly improves weight loss compared to a standard diet plan (Group B).

   The following data was collected:

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

Perform an independent t-test to determine if the new diet plan significantly improves weight loss at a significant level of α = 0.05

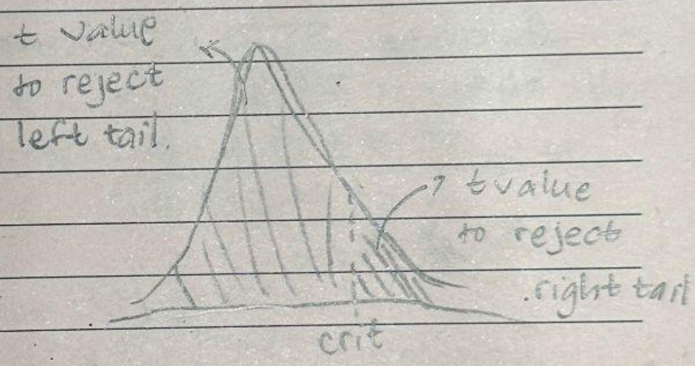
3) $H_0: \mu_{group\,A} = \mu_{group\,B}$ => right tailed

$H_1: \mu_{group\,A} > \mu_{group\,B}$

mean diff = 8 − 6 = 2

SD group A = 2

SD group B = 2.5



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

$$= \frac{2}{\sqrt{\frac{2^2}{25} + \frac{2.5^2}{25}}}$$

$$= 3.13$$

$df = n_A + n_B - 2 = 48$

At $\alpha = 0.05$ (one-tailed) the critical t-value is 1.67722

Since 3.13 > 1.67722, Ho is rejected

EXERCISE 7

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.

1) H_0 : All the means are equal.

H_1 : At least 1 population is different.

Calculate group & overall mean.

A = 15.4

B = 20.4

C = 26

Group mean = 20.6

Calculate SSR / SS between

$SSR = n\Sigma(\bar{x}_j - \bar{\bar{x}})^2$

$= 5(15.4-20.6)^2 + 5(20.4-20.6)^2 + 5(26-20.6)^2$

$= 281.2.$

Calculate SSE / SS within $(\Sigma(x_{ij} - \bar{x}_j)^2)$

$A = (15-15.4)^2 + (16-15.4)^2 + (14-15.4)^2 + (15-15.4)^2 + (17-15.4)^2$

$= 5.2$

$B = (20-20.4)^2 + (22-20.4)^2 + (19-20.4)^2 + (21-20.4)^2 + (20-20.4)^2$

$= 5.2$

$C = (25-26)^2 + (27-26)^2 + (28-26)^2 + (28-26)^2 + (24-26)^2$

$= 10$

SSE = 20.4

Calculate SST.

$SST = SSR + SSE$

$= 301.6$

ANOVA table

Source	Sum of Squares	df	Mean Squares	F
Treatment / between	281.2	2	140.6	
Error / within	20.4	12	1.7	
Total	301.6	14	21.543	82.7

df treatment $= k - 1 = 3 - 1 = 2$, k is no. of groups.

df error $= n - k = 15 - 3 = 12$, n is no. of observation

Mean square treatment $= \frac{SSR}{df\ treat} = \frac{281.2}{2}$

$= 140.6$

Mean square error $= \frac{SSE}{df\ error} = 1.7$

Mean squares total $= \frac{SST}{df\ total} = 21.543$

$F = \frac{MS\ treat}{MS\ error} = 82.7$

Input on calculator F value, df between as numerator, df within as denominator

critical value $= 3.885$

Reject Ho because $F > 3.885$.

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.

2) H_0 : Plant type & fertilizer preference are independent.
H_1 : not.

Expected frequency for every cell.

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

X type A $= \dfrac{40 \times \cancel{30}}{\cancel{90}_{3}} \simeq 13.33$

X type B $= \dfrac{40 \times 35}{90} = 15.56$

X type C $= \dfrac{40 \times 25}{90} = 11.11$

Y type A $= \dfrac{30 \times 30}{90} = 10$

Y type B $= \dfrac{30 \times 35}{90} = 11.67$

Y type C $= \dfrac{30 \times 25}{90} = 8.33$

Z type A $= \dfrac{20 \times 30}{90} = 6.67$

Z type B $= \dfrac{20 \times 35}{90} = 7.78$

Z type C $= \dfrac{20 \times 25}{90} = 5.56$

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

Calculate for every cell:

X, type $A = \frac{(10 - 13.33)^2}{13.33} = 0.83$

X type $B = 1.27 \Rightarrow \frac{(20 - 15.56)^2}{15.56} = 1.27$

X type $C = 0.11$

Y type $A = 2.50$

Y type $B = 0.24$

Y type $C = 1.33$

Z type $A = 0.42$

Z type $B = 0.99$

Z type $C = \underline{3.55} +$

$\chi^2 = 11.24$

Degree of freedom:

df = (no. of rows − 1) × (no. of columns − 1)

= 2 × 2

= 4

From χ^2 distribution table for df = 4 $\alpha = 0.05$

χ^2 crit = 9.448

11.24 > 9.448, H_0 is rejected.

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 α = 0.05

3) Ho programming lang : Mean test score are the same across Python, Java, C++.

Ho study method : —— " —— for study led and instructor.

Ho interaction effect : There is no interaction between lang & method.

Grand mean = 78.944

Group mean =

Python = 85.833

Java = 78.333

C++ = 72.667

Self-study = 74.333

Instructor = 83.556

Python, self = 81.667

Python, instructor = 90.0

Java, self = 73.667

Java, instructor = 83.0

C++, self = 67.667

C++, instructor = 77.667

Sum of Squares Total :

SS total $= (78 - 78.944)^2 + \ldots + (80 - 78.944)^2$

$= 984.944$ (A)

df total $= (n\,p\,q) - 1$

$= (3 \times 3 \times 2) - 1$

$= 17$

n = no. of sample each category

p = no. of language category (A)

q = no of method category (B)

studee 30 lines (6mm spaced)

$MS_{total} = \dfrac{SS_{total}}{df_{total}} = 57.938$

Sum of Squares Between:

$SS_{between} = n[(\text{mean each cat} - \text{grand mean})^2 + \ldots + (\text{mean each cat})^2]$

$= 3[(81.667 - 78.944)^2 + \ldots + (77.667 - 78.944)^2]$

$= 908.2476$

$df_{between} = (p \cdot q) - 1$

$= (3 \times 2) - 1$

$= 5$

$\sigma^2_{between} = \dfrac{SS_{between}}{df_{between}} = 181.650$

$SS_{group\,A} = n \cdot q[(\text{mean every language} - \text{grand mean})^2 + \ldots + (\text{mean every language} - \text{grand mean})^2]$

$= 3 \times 2[(85.833 - 78.944)^2 + (78.333 - 78.944)^2 + (72.667 - 78.944)^2$

$= 523.394$

$df_A = p - 1$

$= 3 - 1$

$= 2$

$MS_A = \dfrac{SS_A}{df_A}$

$= 261.697$

$SSB = np[(\text{mean every method} - \text{grand mean})^2 + \ldots$
$(\text{mean every method} - \text{grand mean})^2]$
$= 3 \times 3[(74.333 - 78.944)^2 + (83.556 - 78.944)^2]$
$= 382.787$

$df_B = q - 1$
$= 2 - 1$
$= 1$

$MS_B = \dfrac{SS_B}{df_B}$
$= 382.787$

$SS_{AB} = SS_{bet} - SS_A - SS_B$
$= 2.0666$

$df_{AB} = (p-1)(q-1)$
$= 2 \times 1$
$= 2$

$MS_{AB} = \dfrac{SS_{AB}}{df_{AB}}$
$= 1.0333$

$SS_{error} = (\text{every sample} - \text{mean every cat})^2 + \ldots + (\text{every sample} - \text{mean every cat})^2$
$= (78 - 81.667)^2 + (90 - 90)^2 + \ldots + (80 - 81)^2$
$= 76.667$

$df_{error} = (n-1)pq$
$= (3-1)\,3 \times 2$
$= 12$

$MS_{error} = \dfrac{SS_{error}}{df_{error}}$
$= 6.389$

$$F_A = \frac{MS_A}{MS_{err}} = 40.961$$

$$F_B = \frac{MS_B}{MS_{err}} = 59.914$$

$$F_{AB} = \frac{MS_{AB}}{MS_{err}} = 0.162$$

p value $A < 0.001$ } Reject H_0 bcs p value < 0.05
p value $B < 0.001$
p value $AB = 0.849$ => Fail to reject bcs p value > 0.05.