Basic Statistics

3rd Semester

—Louisa Mandy Halim

Statistics

Week 1

Three types of data

- Quantitative , consists of numbers weights, heights
- Ranked , —"— relative standing within a group
- qualitative , describing how they feel, look, wear

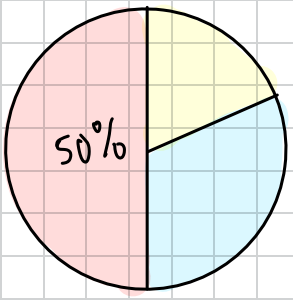
Level of measurement

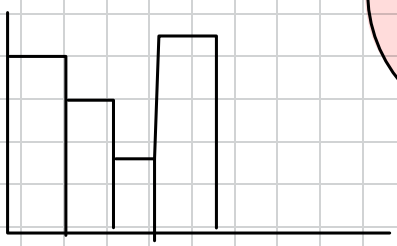
- Sorting observations into different classes or categories
- classyfying

independent & dependent variable

Week 2

- pie chart
- Bar chart





Comparing distributions

example

a bar chart of the number of people playing different card games on wednesday and Thursday

Make simple and easy to understand designs for charts

Stem and leaf

22, 24, 26, 29
13, 14, 15, 16, 17, 19
6, 7

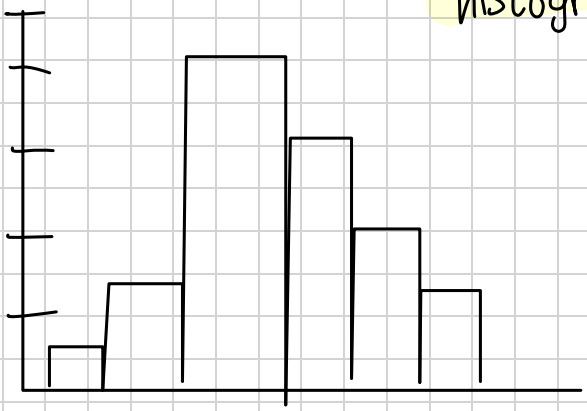
stem	leaf
2	2, 4, 6, 9
1	3, 4, 5, 6, 7, 9
0	6, 7

Histogram

graphical method displaying the shape of distribution

- frequency table

histogram

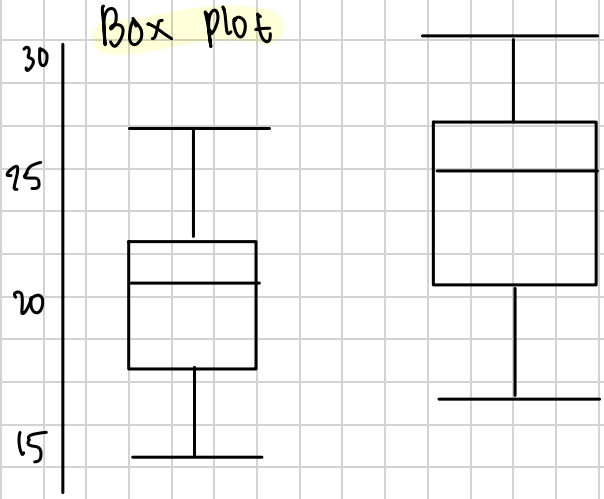


frequency polygons

lower	upper	count

good choice to display frequency distribution

Box plot



- comparing genders
- identifying outliers and for comparing distributions

percentile

80% of the people are shorter than you

No Cap

Exercise 01 — Sep 11, 2024

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

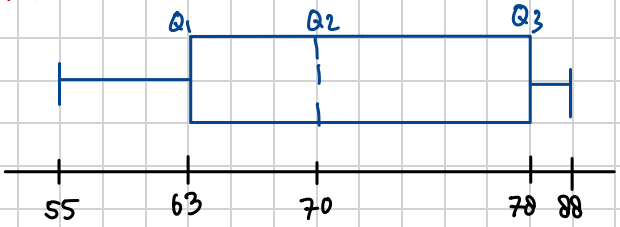
Stem	leaf
6	2, 5, 8
7	0, 3, 5, 5, 8
8	1, 3, 4, 5, 7, 9
9	2, 5, 6, 8
10	0

② Construct a box plot

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

a) Determine the five number summary
- minimum value = 55
- maximum value = 88
- $Q_1 = 63$
- $Q_2 = 70$ (middle)
- $Q_3 = 78$

b) Draw the boxplot



c) identify potential outliers

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

1.5 × 15 = 22.5

upper fence $= Q_3 + (1.5 \times IQR)$
$= 78 + (1.5 \times 15)$
$= 78 + 22.5$
$= 100.5$

lower fence $= Q_1 - (1.5 \times IQR)$
$= 63 - (1.5 \times 15)$
$= 63 - 22.5$
$= 41.5$

Since all data points lie within range 41.5 to 100.5. There are no outliers

Exercise 02 - Sep 25, 2024

① calculate the Trimean for a dataset below

Dataset:

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

Since the dataset is sorted already - find the median:

$Q_1 = 18$
$Q_2 = 30$
$Q_3 = 42$

The trimean for this dataset is 30

Calculate the Trimean

$$\text{Trimean} = \frac{Q_1 + 2 \times \text{Median} + Q_3}{4}$$

$$= \frac{18 + 2(30) + 42}{4}$$

$$= 120/4 = 30$$

Trimean formula:

$$\frac{q_1 + (m \times 2) + q_3}{3}$$

② Geometric Mean

Suppose that the population of a city changes over four years, with the following annual growth rates:

Year 1: + 5% → 1.05
year 2: + 10% → 1.1
Year 3: - 3% → 0.97
Year 4: + 6% → 1.06

Geometric mean formula:

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots}$$

Calculate the geometric mean of the growth rates to find the average population growth rate over these 4 years

$$GM = \sqrt[4]{1.05 \cdot 1.1 \cdot 0.97 \cdot 1.06}$$

$$= 1.0426$$

$$= 1.0426 - 1 = 0.0426 \times 100$$

$$= 4.26\% \rightarrow \text{the growth rate over 4 years}$$

③ Trimmed Mean

10% of 10 = $\frac{10}{100} \times 10 = 1$, remove 1 from each end

10

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

calculate the 20% trimmed mean (Trim 10% from both sides)

$$\frac{70+72+75+80+85+90+95}{8} = 82{,}375$$

add Conclusion

□ More examples

a figure skating competition produces the following scores:

6.0 ; 8.1 ; 8.3 ; 9.1 ; 9.9

$$\text{mean} = \frac{6.0+8.1+8.3+9.1+9.9}{5}$$

$$= \frac{41.4}{5} \rightarrow 8.28$$

Trim the mean by 40%

- remove the lowest 20% and the highest 20% of values, eliminating the scores: 6.0 & 9.9

$$\frac{8.1+8.3+9.1}{3} = 8.5$$

a mean trimmed by 40% would equal 8.5 vs 8.28, which reduced the original mean by 0.22 points

Exercise 03 - Oct 02, 2024

① You have 8 people, and you need to select and arrange 4 of them in a row of photo. How many different ways can you arrange?

$n = 8$, $r = 4$ | $\frac{n!}{(n-r)!} = \frac{8!}{4!} = \frac{40.320}{24} \rightarrow$ 1,680 ways

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

$n = 7$, $r = 4$ | $nCr = \frac{n!}{(n-r)!\,r!}$ | $\frac{7!}{3! \times 4!} = \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{3! \times 4!}$

$= \frac{210}{6} = 35$ ways

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

Combination formula : $nCr = \frac{n!}{(n-r)! \cdot r!}$

Hypergeometric formula : $P = \frac{k C x \; (N-k) \; C \, (n-x)}{(n-r)! \cdot r!}$

$15C_2 = \frac{15!}{13! \cdot 2!} = \frac{15 \times 14}{2} = 105$

$10C_3 = \frac{10!}{7! \times 3!} = \frac{10 \times 9 \times 8}{6}$ (→ 120)

$25C_5 = \frac{25!}{20! \times 5!} = \frac{25 \times 24 \times 23 \times 22 \times 21}{120} \rightarrow 53.130$

Probability → $\frac{10C_3 \times 15C_2}{25C_5} = \frac{120 \times 105}{53.130} = 0,12372$

- more examples

in how many ways a committee consisting of 5 men and 3 Women, can be chosen from 9 men and 12 women

→ $9C_5 = \frac{9!}{4! \cdot 5!} \rightarrow \frac{9 \times 8 \times 7 \times 6}{4!} =$ 126 ways

→ $12C_3 = \frac{12!}{9!\,3!} \rightarrow \frac{12 \times 11 \times 10}{6} =$ 220 ways

Total number of ways

↳ $126 \times 220 = 27720$

Exercise 04- October 9, 2024

① find the percentage returns from an investment over 5 consecutive years, were:

Year 1: 10% → 0,1
Year 2: 15% → 0,15
Year 3: -5% → 0,95
Year 4: 8 % → 0,8
Year 5: 12% → 0,12

$GM = \sqrt[5]{0,1 \times 0,15 \times 0,95 \times 0,8 \times 0,12}$

$= 1,073$

$1,073 - 1 = 0,073 \times 100$

$= 7,3\ \%$

② Box plot

Dataset:

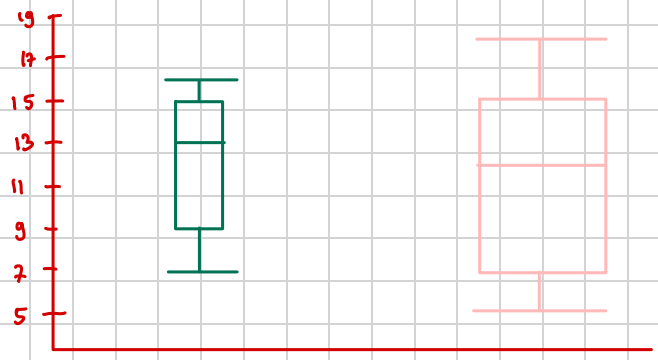
Group A: 7, 9, 12, 13, 14, 15, 16
Group B: 5, 7, 8, 10, 12, 15, 18

a) calculate the five number summary

- min value = 7
- max value = 16
- Q_1 = 9
- Q_2 = 13
- Q_3 = 15

- min value = 5
- max value = 18
- Q_1 = 7
- Q_2 = 10
- Q_3 = 15

b)



c) Compare the distributions of two groups based on the box plots.

- which group has a higher median? – Group A
- Are there any outliers?
  no outliers

(3) a card is drawn from a standard deck of 52 cards, and then a coin is flipped. what is the probability of drawing "king" from the deck and flipping a "Tail"?

1 Deck = 52 cards
kings = 4

$$\frac{4}{52} \times \frac{1}{2} = \frac{1}{26}$$

↑ four kings out of the total cards

↘ tail side, out of the 2 sides (coin)

(4) Stem and leaf

Department x : 12, 14, 17, 19, 21, 24, 26, 28, 30, 32
Department y : 13, 16, 18, 20, 23, 25, 27, 29, 31, 33

Back to back stem and leaf display

Department x		Department y
2, 4, 7, 9	1	3, 6, 8
1, 4, 6, 8	2	0, 3, 5, 7, 9
0, 2	3	1, 3

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

$$5C_3 = \frac{5!}{2! \cdot 3!} = \frac{20}{2} = 10 \text{ ways}$$

$$\frac{10}{32} = \frac{5}{16}$$ Probability

after 5 flips → $2^5 = 32$

Using the formula

$n = 5$ (number of trials)
$x = 3$ (number of successes)
$\pi = 0.5$ (probability of success)

$$\frac{n!}{x!(n-x)!} \cdot \pi^x (1-\pi)^{n-x}$$

$$= 10 \times 0.5^3 \times 0.5^2$$

$$= 10 \times \frac{5}{10} \times \frac{5}{10} \times \frac{5}{10} \times \frac{5}{10} \times \frac{5}{10}$$

$$= \frac{5}{16} \rightarrow \text{Probability}$$

⑤ 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?.

formula: $\binom{n}{x} \cdot p^x \cdot q^{n-x} \rightarrow \frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x}$

$n = 15$
$x = 12, 13, 14, 15$
$p = 0,8$
$q = 0,2$

- for $15C_{12}$
  - $\frac{15!}{3! \cdot 12!} \times 0,8^{12} \times 0,2^3 = 0,227$
- for $15C_{13}$
  - $\frac{15!}{2! \cdot 13!} \times 0,8^{13} \times 0,2^2 = 0,236$
- for $15C_{14}$
  - $\frac{15!}{1! \cdot 14!} \times 0,8^{14} \times 0,2^1 = 0,137$
- for $15C_{15}$
  - $\frac{15!}{15!} \times 0,8^{15} \times 0,2^0 = 0,035$

Total = 0,635

the probability

(7)

Hours of sunlight x	Height (cm) y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
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			

Calculate the Pearson correlation coefficient

mean of x = 2, 4, 6, 8, 10

mean of y = 10, 15, 20, 25, 30

formula : $$\frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2 \cdot \Sigma (y-\bar{y})^2}} = \frac{100}{\sqrt{1000}} = \frac{100}{100} \rightsquigarrow 1$$

Exercise 05 - October 23 Oct, 2024

(1) Find the standard deviation from those data

Scores: 70, 85, 75, 90, 88

mean → $\frac{70+85+75+90+88}{5} = 82,2$

$$\frac{(70-82,5)^2 + (85-82,5)^2 + (75-82,5)^2 + (90-82,5)^2 + (88-82,5)^2}{5} = \frac{(-12,2)^2 + (2,8)^2 + (-4,2)^2 + (7,8)^2 + (5,8)^2}{5}$$

variance = 53,76

Standard deviation = $\sqrt{53,76}$ → 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 z-table

$n = 100$
$p = 30\% \rightarrow 0.3$
$q = 0.70$

$M = n \times p$
$= 100 \times \frac{3}{10} = 30$

$\sigma = \sqrt{100 \times 0.3 \times 0.7}$
$= \sqrt{21} \sim 4.58$

because $25 \rightarrow 25 - 0.5 = 24.5$

$z = \frac{24.5 - 30}{4.58} = \frac{-5.5}{4.58} \rightarrow -1.20$

from z table, $-1.20 \rightarrow 0.1151$

The probability that fewer than 25 people prefer coffee is approximately 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

$n = 100$
$p = 0.4$
$q = 0.6$

$M = n \times p$
$= 100 \times 0.4$
$= 40$

$\sigma = \sqrt{100 \times 0.4 \times 0.6}$
$= 2\sqrt{6} \rightarrow 4.89$

$z = \frac{44.5 - 40}{4.89} \rightarrow \frac{4.5}{4.89} = 0.92$

from z table $\rightarrow$ 0.8212

$1 - 0.8212 = 0.1788 \rightarrow 17.88\%$

Exercise 06 — November 20, 2024

① a company claims their light bulb last 1000 hours on average.

950, 960, 970, 980, 1020, 1030, 990, 1010, 1000, 955 → mean = 990,5

Test whether the mean lifespan differs significantly from 1000 hours using $\alpha = 0{,}05$

formula t statistic : $t = \frac{m - \mu}{s/\sqrt{n}}$

$s = 25{,}87$
mean = 990,5
$\mu = 1000$

$= \frac{990{,}5 - 1000}{25{,}87/\sqrt{10}} = -1{,}16$

Compare t statistic with Critical Value

Degrees of freedom : n−1 → 10−1 = 9

at a = 0,05 (two tailed)
the critical t value is approximately ± 2.262 (from t table)

Since t = −1,16 falls within the range [−2.262 ; 2.262], we fail to reject the null hypothesis

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 $\alpha = 0.05$

Solution:
Null Hypothesis (H_0): The training program has no effect on weight ($\mu_d = 0$).
Alternative Hypothesis (H_1): The training program reduces weight ($\mu_d < 0$).

Sd or standard deviation formula:

$Sd = \sqrt{\frac{\Sigma (di - \bar{d})^2}{n-1}} = 0{,}835$

Conclusion: ($\alpha = 0{,}05$)
Degrees of freedom = 8−1 → 7
t crit = 2,365

Since −10,59 < −2,365
reject null hypothesis
significant difference

formula t statistic : $\frac{\bar{d}}{sd/\sqrt{n}}$ → $-3{,}125 / 0{,}835/\sqrt{8} = -10{,}59$

Standard deviation = $\frac{\Sigma d}{n}$ → $\frac{-3-3-5-2-3-3-3-3}{8} = -3{,}125$

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 $\alpha = 0.05$

calculate t statistic

$$t = \frac{x_A - x_B}{\sqrt{\frac{SA^2}{nA} + \frac{SB^2}{nB}}}$$

$$t = \frac{8-6}{\sqrt{\frac{2^2}{25} + \frac{2.5^2}{25}}} \rightarrow \frac{2}{\sqrt{0.16 + 0.125}} \rightarrow \frac{2}{0.64} \rightarrow 3.13$$

Degrees of freedom (df) = $n_A + n_B - 2 = 48$

from t table (one tail) = 1.679

since t value > critical t value

3.13 > 1.679

reject null hypothesis

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

mean 1 | mean 2 | mean 3

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.

$n_i = 5$, $I = 3$

total mean

SSTR = $\sum(15.4 - 20.6)^2$

SSE = each

$5\left(\left(15 - 15.4\right)^2 + (16\right.$

SSTR → $\sum$(mean 1 − total mean)^2 + $\sum$(mean 2 − ... + ...

$5($ ...

- Fertilizer A mean

$$\frac{15+16+14+15+17}{5} = \frac{77}{5} \rightarrow 15{,}4 \quad (\bar{x}A)$$

- Fertilizer B mean

$$\frac{20+22+19+21+20}{5} = \frac{102}{5} \rightarrow 20{,}4 \quad (\bar{x}B)$$

- Fertilizer C mean

$$\frac{25+27+26+28+24}{5} = 26 \quad (\bar{x}C)$$

- Overall means

$$\bar{x} = \frac{20+22+19+21+20+25+27+26+28+4}{15} = 20{,}6 \quad (\bar{x})$$

Sum of Squares

- Sum of squares (SST)

$SST = (15-20{,}6)^2 + (16-20{,}6)^2 + \ldots + (24-20{,}6)^2 = 89{,}2$

- Sum of Squares between group (SSB)

$$SSB = 5\left[(15{,}4-20{,}6)^2 + (20{,}4-20{,}6)^2 + (26-20{,}6)^2\right]$$
$$= 281{,}04$$

- Sum of Squares within group (SSW)

$SSW = SS\ total + SS\ between$

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 fertilizer has a significant effect on plant growth.

→ The F-static ($F = 24{,}35$) is 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 $\alpha = 0.05$.

Calculate the expected Frequencies

formula = $\frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$

Fertilizer X

- plant type A: $\frac{30 \times 40}{90} = \frac{1200}{90} \rightarrow 13,33$
- plant type B: $\frac{35 \times 40}{90} = \frac{140}{9} \rightarrow 15,56 / 15,55$
- plant type C: $\frac{25 \times 40}{90} = \frac{100}{9} \rightarrow 11,11$

Fertilizer Y

- plant type A: $\frac{30 \times 30}{300} = 10$
- plant type B: $\frac{35 \times 30}{300} = \frac{35}{3} \rightarrow 11,66$
- plant type C: $\frac{25 \times 30}{90\ 3} = \frac{25}{3} \rightarrow 8,33$

Fertilizer Z

- plant type A: $\frac{30 \times 20}{90\ 3} = \frac{20}{3} = 6,66$
- plant type B: $\frac{35 \times 20}{90} = \frac{70}{9} \rightarrow 7,77$
- plant type C: $\frac{25 \times 20}{90} = \frac{50}{9} \rightarrow 5,55$

Fertilizer	Plant A	Plant B	Plant C	total
X	13,33	15,56	11,11	40
y	10	11,66	8,33	30
Z	6,66	7,77	5,55	20
Total	29,99 → 30	34,99 → 35	24,99 → 25	

Compute chi-Square statistic:

$$x^2 = \sum \frac{(Oij - Eij)^2}{Eij}$$

Fertilizer X

- plant type A: $\frac{(10 - 13,33)^2}{13,33} = 0,83$
- plant type B $\frac{(20 - 15,56)^2}{15,56} = 1,26$
- plant type C $\frac{(10 - 11,11)^2}{11,11} = 0,11$

Fertilizer y:

- plant type A: $\frac{(15 - 10)^2}{10} = 2,5$
- plant type B: $\frac{(10 - 11,66)^2}{11,66} = 0,23$
- plant type C: $\frac{(5 - 8,33)^2}{8,33} = 1,33$

Fertilizer Z

- Plant type A $\frac{(5 - 6,66)^2}{6,66} = 0,41$
- Plant type B $\frac{(5 - 7,77)^2}{7,77} = 0,98$
- Plant type C $\frac{(10 - 5,55)^2}{5,55} = 3,56$

Summing all given values = 11,21

Degrees of freedom

df = (number of rows − 1) × (number of Columns − 1)
= (3−1) × (3−1)
= 2 × 2
df = 4

Determine Critical value

from Chi square table

for df = 4, $\alpha = 0.05$

the critical value is 9,488

compare test statistics to critical value

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

Here:

$x^2 = 11,21$

$x^2_{critical} = 9,488$

Since $11,21 > 9,488$

reject the null hypothesis

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

- find Grand / Group mean

Python Self Study = $\frac{78+82+85}{3} = 81,66$

Java Self Study = $\frac{72+75+74}{3} = 73,66$

C++ Self Study = $\frac{65+68+70}{3} = 67,66$

Python instructor-led

$\frac{90+88+92}{3} = 90$

C++ $\frac{78+75+80}{3} = 77,66$

Java instructor

$\frac{85+80+84}{3} = 83$

Grand mean ($\bar{x}$) = $\frac{81,66+73,66+67,66+90+83+77,66}{6} = 78,94$

Group means

Python	(81,66 + 90) ÷ 2 = 85,83
Java	(73,66 + 83) ÷ 2 = 78,33
C++	(67,66 + 77,66) ÷ 2 = 72,66
Self	total all ÷ 9 = 74,33
instructor	— " — = 83,55

Python & self study	81,66
python & instructor	90
Java and Self	73,66
Java & instructor	83
C++ & self	67,66
C++ & instructor	77,66

Compute Sum of Squares:

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

$\bar{x} = 78.94$

- Sum of Squares for programming language (A)

$$SSA = 6 \cdot (85.83 - 78.94)^2 + 6 \cdot (78.33 - 78.94)^2 + 6 \cdot (72.66 - 78.94)^2$$

Python Java C++

$SSA = 523.6956$

- Sum of squares for factor study method (B)

$$SSB = 9\,(74.33 - 78.94)^2 + 9\,(83.55 - 78.94)^2$$

self study instructor

$SSB = 382.5378$

- Sum of Squares Within (error)

Single numbers from tables - Group means

- SS python & self study = $(78 - 81.66)^2 + (82 - 81.66)^2 + (85 - 81.66)^2 = 24.6668$
- SS python & instructor-led = $(90 - 90)^2 + (88 - 90)^2 + (92 - 90)^2 = 8$
- SS Java & self study = $(72 - 73.66)^2 + (75 - 73.66)^2 + (74 - 73.66)^2 = 4.6668$
- SS Java & instructor-led = $(85 - 83)^2 + (80 - 83)^2 + (84 - 83)^2 = 14$
- SS C++ & self study = $(65 - 67.66)^2 + (68 - 67.66)^2 + (70 - 67.66)^2 = 12.9512$
- SS C++ & instructor-led = $(78 - 77.66)^2 + (75 - 77.66)^2 + (80 - 77.66)^2 = 12.6668$

$SSE = 24.6668 + 8 + 4.6668 + 14 + 12.9512 + 12.6668$

$SSE = 76.9528$

- Total Sum of Squares

all Sigle numbers from table - $\bar{x}$

- SS Total = $(78 - 78.94)^2 + (82 - 78.94)^2 + (85 - 78.94)^2 + (72 - 78.94)^2 + (75 - 78.94)^2 + (74 - 78.94)^2 + (65 - 78.94)^2 + (68 - 78.94)^2 + (70 - 78.94)^2 + (90 - 78.94)^2 + (88 - 78.94)^2 + (92 - 78.94)^2 + (85 - 78.94)^2 + (80 - 78.94)^2 + (84 - 78.94)^2 + (78 - 78.94)^2 + (75 - 78.94)^2 + (80 - 78.94)^2 = 984.9444$

SS interaction = SS Total − SSA − SSB − SSE

= 984,9444 − 523,6956 − 382,5378 − 76,9528

= 1,7582

Degrees of freedom

df A = 2

df B = 1

df interaction = 2

df within = 12

df total = 17

Mean Squared and FF- statistics

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			

Week 4

Probability of a single event

$$\text{Probability} = \frac{\text{Possible outcomes}}{\text{total outcomes}}$$

- probability to shuffle an ace in a one deck card
- the chance of dices
- probability of two (or more) independent events
- flipping a coin

Exercise 1:
Let's consider a scenario where you're trying to predict the exam score based on the number of hours studied, but the sample data has more variability due to other factors like individual learning styles, distractions, or even exam difficulty.

Hours of Study (X)	Exam Score (Y)
1	52
2	59
3	62
4	64
5	72
6	80
7	74
8	83
9	91
10	89

1. Find the regression equation (Y' = bX + A).
2. After that, predict the exam score for someone who studied for 7 hours. Compare the results with the data in the table, explaining why the predictions are not the same.
3. Then, predict the exam score for someone who studied for 11 hours.

Exercise 2:
You are given the following sample data showing the height (in inches) and weight (in pounds) of 5 people:

Height (X)	Weight (Y)
60	140
62	145
64	160
66	170
68	155

1. Find the regression equation (Y' = bX + A).
2. After that, predict the weight for someone who is 70 inches tall.

Exercise 1

Louisa Mandy Halim - 2702325552

1)

Hours of Study (X)	Exam Score (Y)
1	52
2	59
3	62
4	64
5	72
6	80
7	74
8	83
9	91
10	89

$x \cdot y$	x^2	y^2
52	1	2.704
118	4	3.481
186	9	3.844
256	16	4.096
360	25	5.184
480	36	6400
518	49	5476
664	64	6889
819	81	8281
890	100	7921

$\Sigma = 55 \quad \Sigma = 726 \quad \Sigma = 4343 \quad \Sigma = 385 \quad \Sigma = 54.276$

(1) $b_0 = \dfrac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{(n(\Sigma x^2) - (\Sigma x)^2)}$

$b_0 = \dfrac{(726 \times 385) - (55 \times 4343)}{(10 \times 55) - (55)^2}$

$b_0 = \dfrac{279.510 - 238.865}{825}$

$b_0 = \dfrac{40.645}{825} \Rightarrow 49,3$

$b_1 = \dfrac{(n(\Sigma xy) - (\Sigma x)(\Sigma y))}{n(\Sigma x^2) - (\Sigma x^2)}$

$b_1 = \dfrac{(10 \times 4343) - (55 \times 726)}{(10 \times 385) - (55)^2}$

$b_1 = \dfrac{43430 - 39930}{3850 - 3025} \Rightarrow 4,24$

$y' = b_0 + b_1 x$

$y' = bx + A$

$y' = 4,24x + 49,3$

(2) to predict the student who studied for 7 hours

$y' = 4,24x + 49,3$

$x = 7 \rightarrow 7 \text{ hours}$

$y' = 4,24(7) + 49,3$

$y' = 78,98$

The regression equation captures the general trend and minimizes errors, but may not precisely align with each individual data point.

(3) Predict after studying for 11 hours

$y' = 4,24x + 49,3$

$x = 11 \rightarrow \text{hours}$

$y' = 4,24(11) + 49,3$

$y' = 95,94$

Exercise 2

①

	x	y	x^2	x.y
	60	140	3600	8400
	62	145	3844	8990
	64	160	4096	10.240
	66	170	4356	11.220
	68	155	4624	10.540
$\Sigma =$	320	770	20.520	49.390

$$b_0 = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$b_0 = \frac{5(49.390) - (320)(770)}{5(20.520) - (320)^2}$$

$$b_0 = \frac{550}{200} \Rightarrow 2,75$$

$$b_1 = \frac{(\Sigma y \cdot \Sigma x^2) - (\Sigma x \cdot \Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_1 = \frac{(770 \times 20.520) - (320 \times 49.390)}{200}$$

$$b_1 = -22$$

$y' = bx + A$

$y' = b_0 + b_1 x$

$y' = 2,75x - 22$

② $y' = 2,75(70) - 22$

$y' = 170,5$