

ALL STATISTICS EXERCISES

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

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 $1.5 \times$ H-spread to identify outliers)
- c. Identify any potential outliers (outside value or/and far out value)

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

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%

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

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

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%

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
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”?
 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.

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

EXERCISE 5

1. Find the standard deviation from these data

Scores: 70, 85, 78, 90, 88

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!
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.

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 $\alpha = 0.05$
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 $\alpha = 0.05$

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 $\alpha = 0.05$

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 $\alpha = 0.05$)

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

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 $\alpha = 0.05$.

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 $\alpha = 0.05$