Exercise 1

A study has been conducted on the recovery time (in days) of a sample of patients with the flu who **did not receive any treatment**, obtaining the following data:

Days: x_i	**n_i**	**f_i**	**F_i**
[13, 15)	25	0.3125	0.3125
[15, 17)	35		
[17, 19)	17	0.2125	0.9625
[19, 21)	2		0.9875
[21, 23)	1	0.0125	1
	80		

a) **Complete** the table.

b) Are there any **outliers**? If so, in which **interval** would they be? Justify your answer by constructing a **box-and-whisker plot**.

c) It is determined that a patient who takes **18 days or more** to recover should undergo an urgent check-up to assess hospitalization. What **percentage** of the sample should undergo this check-up?

d) Considering its **skewness** and **kurtosis**, could we justify that this sample comes from a normal distribution? Describe the shape of the sample.

e) **A treatment reduces the recovery time** to half the days plus one (y = x/2 + 1). Which **mean** would **be more representative**, the recovery time without treatment or with treatment?

Use the following sums for the calculations:

$\sum n_i =$	80
$\sum x_i\, n_i =$	1278
$\sum x_i^2\, n_i =$	20652
$\sum (x_i - \overline{x})^2 n_i =$	235.95
$\sum (x_i - \overline{x})^3 n_i =$	297.6975
$\sum (x_i - \overline{x})^4 n_i =$	2508.8849

Exercise 2

At Hogwarts, Professor Severus Snape needs to prepare a potion whose main ingredient is Mandrake to cure a group of petrified students. To know when the Mandrake will be available for the potion, he asks Professor Pomona Sprout to tell him if there is a linear relationship between the number of days the plant takes to germinate (Y) and the ambient temperature (X), having observed the following data in previous crops:

X: Temperature (ºC)	31	25	21	17
Y: Time (days)	2	4	5	8

Sums	
$\sum X_i n_i$	94
$\sum Y_i n_i$	19
$\sum X_i^2 n_i$	2316
$\sum Y_i^2 n_i$	109
$\sum X_i \, Y_i \, n_i$	403

a) Calculate the **statistics** for the following table, specifying the **units** in each case.

Statistic	Value	Unit
$\overline{x}$		
$\overline{y}$		
S_x^2		
S_y^2		
S_{xy}		
r^2		

b) Use a linear regression model to **predict the number of days** it will take for a plant to germinate if the ambient temperature is 12ºC.

c) Extract the slop of the previous linear regression model and **interpret** how the plant's **germination time will change for each degree** the temperature drops.

d) Use the appropriate linear regression model to **predict what ambient temperature** would be needed for the plant to germinate in 3 days.

e) Are all the previous **predictions equally reliable**? Explain why.

Exercise 3

A treatment for myopia causes dizziness and eye irritation as side effects. Studies indicate that 8% of patients experience dizziness, while 15% experience irritation. It has been found that the probability of experiencing dizziness **knowing they experience irritation** is 25%. Calculate:

a) The probability that a patient experiences **at least** one symptom.
b) The probability that a patient experiences **only** irritation.
c) The probability that a patient experiences dizziness **knowing they do not experience irritation**.
d) Can dizziness and irritation be considered **independent symptoms**?

Exercise 4

A test for COVID-19 detection has a sensitivity of 98% and a specificity of 95%. The prevalence of COVID-19 in the population is 0.7%.

a) **If a person tests positive**, what is the probability that they have COVID-19? How is this probability called?
b) **If a person tests negative**, what is the probability that they do not have COVID-19? How is this probability called?
c) Calculate the probability of a **correct diagnosis** provided by the test.

Exercise 5

A second test for COVID-19 detection has a **probability of correct diagnoses** of 0.97.

a) In a group of 10 people, calculate the probability that it **correctly diagnoses** at least 8 of them.
b) In a group of 200 people, calculate the probability that it **fails in diagnosing** at most 4 of them.

Exercise 6

The weight of a specie of rat is studied, and it is found to follow a normal (Gaussian) distribution with **unknown mean** and standard deviation $\sigma = 100$ g.

a) Knowing that the probability that a rat weighs 368 g or less is 0.75, **calculate the mean** of the distribution.
b) What is the **probability that a rat weighs more** than 444 g?
c) If a random rat is selected, what is the **probability that it weighs between** 172 g and 444 g?