

EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

Pharmacy/Biotechnology 1st year

Version B

February 4, 2022

Name:

DNI:

Group:

Duration: 1 hour.

- (5 pts.) 1. The account of the pharmacist-tiktoker @farmaceutikofer has numerous informative-satirical videos on infectious diseases, sex education and healthy habits. The duration, in seconds, of his last 30 videos is as follows:

31	32	33	33	35	35	37	37	38	39	39	41	41	42	45
46	46	47	48	48	48	52	52	54	55	55	56	57	59	61

- (a) Compute the average and the most frequent value of the duration of the videos.
- (b) How is the relative dispersion of the sample?
- (c) Construct the frequency table of the sample grouping the data into classes of amplitude 5 seconds and plot the histogram.
- (d) Compute the interquartile range and explain its meaning.
- (e) Has the distribution a normal kurtosis?
- (f) The ticktoker @pauloGshow uploads satirical videos about high school teachers with a mean duration of 45 seconds and a variance of 91 seconds. If both @farmaceutikofer and @pauloGshow upload a 51-second video, which of the two tiktokers has uploaded a shorter video in relation to their own account?

Use the following sums for the computations:

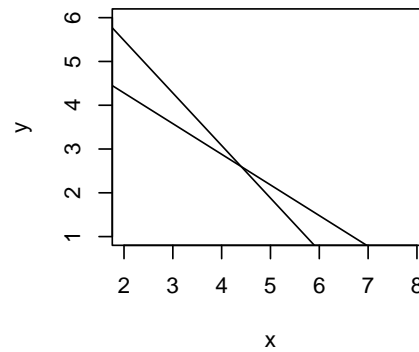
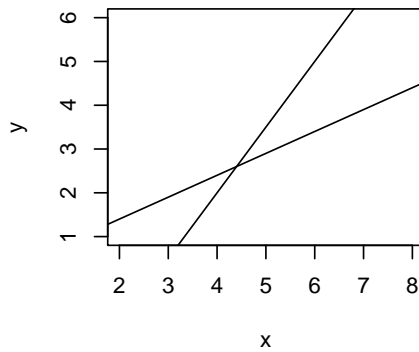
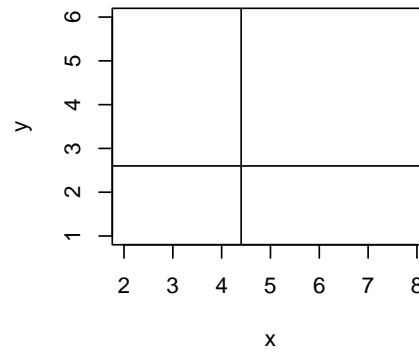
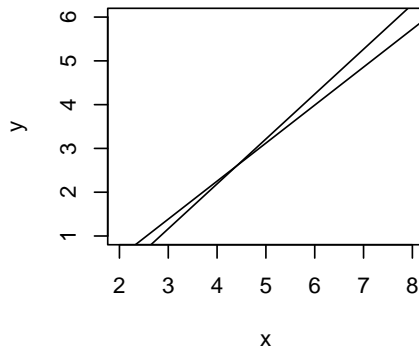
$$\sum x_i = 1342 \text{ s}, \sum x_i^2 = 62302 \text{ s}^2, \sum (x_i - \bar{x})^3 = 2870.46 \text{ s}^3 \text{ y } \sum (x_i - \bar{x})^4 = 315273.2 \text{ s}^4.$$

- (5 pts.) 2. The Regional Ministry of Health of the Community of Madrid realizes a possible relationship between the level of air pollution and the number of cases of pneumonia in the population in the first 10 weeks of the year. To verify this, the variable X registers the number of pollution meters that exceed the pollution limits each week, and the variable Y indicates the number of people affected by pneumonia in each week.

X	3	3	5	6	7	8	3	4	2	3
Y	2	1	2	3	6	6	2	2	1	1

- (a) Are the number of people affected by pneumonia and the number of meters that exceed the pollution limits two linearly independent variables?
- (b) How does the number of people affected by pneumonia change in relation to the number of meters that exceed the pollution limits?
- (c) Justify whether or not the linear relationship between the two variables is well explained and in what proportion.
- (d) In the week following the construction of the linear regression model of Y on X , 5 meters are expected to exceed the pollution limits. How many people are expected to be affected by pneumonia in that week? Is this prediction reliable?

(e) Which of the following diagrams best represents the regression lines?



Use the following sums for the computations:

$\sum x_i = 44$ meters, $\sum y_i = 26$ persons, $\sum x_i^2 = 230$ meters², $\sum y_i^2 = 100$ persons² y $\sum x_i y_i = 146$ meters·persons.