

## EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy

Version A

March, 26 2019

**Duration:** 1 hour and 15 minutes.

- (5 pts.) 1. The time required by a drug  $A$  to be effective has been measured in a sample of 150 patients. The table below summarizes the results.

Response time	Patients
(0, 5]	5
(5, 10]	15
(10, 15]	32
(15, 20]	36
(20, 30]	42
(30, 60]	20

- Are there outliers in the sample? Justify the answer.
- What is the minimum time for the 20% of patients with highest response time?
- What is the average response time? Is the mean representative? What is the average response time of drug  $B$ ? Is more or less representative than the one of drug  $A$ ?
- Can we assume that the sample comes from a normal population?
- If we take another sample of patients with mean 18 min and standard deviation 15 min, in which group is greater a response time of 25 min?

Use the following sums for the computations:  $\sum x_i = 3105$  min,  $\sum x_i^2 = 83650$  min<sup>2</sup>,  $\sum (x_i - \bar{x})^3 = 206851.65$  min<sup>3</sup> y  $\sum (x_i - \bar{x})^4 = 8140374.96$  min<sup>4</sup>.

### Solution

- $Q_1 = 12.5$  min,  $Q_2 = 17.5$  min,  $IQR = 12.5$  min,  $f_1 = -6.25$  min and  $f_2 = 43.75$  min. Therefore there are outliers in the sample since the upper limit of the last interval is 60.
- $P_{80} = 27.619$  min.
- $\bar{x} = 20.7$  min,  $s^2 = 129.1767$  min<sup>2</sup>,  $s = 11.3656$  min and  $cv = 0.5491$ . The mean is not very representative since the  $cv > 0.5$ .
- Let  $Y$  be the response time for drug  $B$ .  $\bar{y} = 155.5$ ,  $s_y = 5.6828$  and  $cv_y = 0.3702$ , therefore the mean of  $B$  is more representative since the coef. of variation is smaller.
- For the first sample  $z_i = 2.578$  and for the second one  $z_i = 0.4667$ , thus a time of 25 min is relatively greater in the second sample.

- (1.5 pts.) 2. In a regression study about the relation between two variables  $X$  and  $Y$  we got  $\bar{x} = 7$  and  $r^2 = 0.9$ . If the equation of the regression line of  $Y$  on  $X$  is  $y - x = 1$ , compute

- The mean of  $Y$ .
- The equation of the regression line of  $X$  on  $Y$ .
- ¿What value does this regression model predict for  $x = 6$ ? ¿And for  $y = 10$ ?

---

**Solution**


---

- (3.5 pts.) 3. In a tennis club the age ( $X$ ) and the height ( $Y$ ) of the ten players conforming the female youth team has been measured.

Age (years)	9	10	11	12	13	14	15	16	17	18
Height (cm)	128	144	148	154	158	161	165	164	166	167

- (a) Plot the scatter plot.  
 (b) Which regression model best fits these data, the linear or the logarithmic?  
 (c) ¿What is the expected height of a player 12.5 years old according to the best of two previous models?

Use the following sums for the computations:

$$\begin{aligned} \sum x_i &= 135 \text{ years}, \sum \log(x_i) = 25.7908 \log(\text{years}), \sum y_j = 1555 \text{ cm}, \sum \log(y_j) = 50.4358 \log(\text{cm}), \\ \sum x_i^2 &= 1905 \text{ years}^2, \sum \log(x_i)^2 = 67.0001 \log(\text{years})^2, \sum y_j^2 = 243191 \text{ cm}^2, \sum \log(y_j)^2 = 254.4404 \\ &\log(\text{cm})^2, \\ \sum x_i y_j &= 21303 \text{ years} \cdot \text{cm}, \sum x_i \log(y_j) = 682.9473 \text{ years} \cdot \log(\text{cm}), \sum \log(x_i) y_j = 4035.0697 \log(\text{years}) \text{cm}, \\ \sum \log(x_i) \log(y_j) &= 130.2422 \log(\text{years}) \log(\text{cm}). \end{aligned}$$

---

**Solution**


---