

EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy

Version A

May, 27 2019

Name:

DNI:

Group:

Duration: 1 hour and 15 minutes.

- (5 pts.) 1. A study tries to determine the effect of smoking during the pregnancy in the weight of newborns. The table below shows the daily number of cigarettes smoked by mothers (X) and the weight of the newborn (all of them are males) (Y).

Daily num cigarettes	10.00	14.00	8.00	11.00	7.00	6.00	2.00	5.00	9.00	9.00	4.00	6.00
Weight (kg)	2.55	2.44	2.68	2.65	2.71	2.85	3.45	2.93	2.67	2.59	3.02	2.72

- Give the equation of the regression line of the weight of newborns on the daily number of cigarettes and interpret the slope.
- Which regression model is better to predict the weight of newborns, the logarithmic or the exponential?
- Use the best of the two previous regression models to predict the weight of a newborn whose mother smokes 12 cigarettes a day. Is this prediction reliable?

Use the following sums for the computations:

$\sum x_i = 91$ cigarettes, $\sum \log(x_i) = 23.0317 \log(\text{cigarettes})$, $\sum y_j = 33.26$ kg, $\sum \log(y_j) = 12.1857 \log(\text{kg})$,
 $\sum x_i^2 = 809$ cigarettes², $\sum \log(x_i)^2 = 47.196 \log(\text{cigarettes})^2$, $\sum y_j^2 = 92.9708$ kg², $\sum \log(y_j)^2 = 12.4665 \log(\text{kg})^2$,
 $\sum x_i y_j = 243.61$ cigarettes·kg, $\sum x_i \log(y_j) = 89.3984$ cigarettes·log(kg), $\sum \log(x_i) y_j = 62.3428$ log(cigarettes)kg, $\sum \log(x_i) \log(y_j) = 22.8753 \log(\text{cigarettes}) \log(\text{kg})$.

- (5 pts.) 2. The table below summarize the time that took to the runners to reach the finish in a long-distance race in Madrid:

Time (min)	Num runners
(30, 35]	15
(35, 40]	35
(40, 45]	40
(45, 50]	10

In a another race in Paris, the mean of time was 40 minutes, the standard deviation 5 minutes and the coefficient of skewness 0.75.

- What percentage of runners took less than 42 minutes to reach the finish in Madrid?
- Compute and interpret the interquartile range of the time for Madrid race.
- In which race the mean of the time is more representative?
- In which race the time have a more symmetric distribution?
- In which race a time of 39 minutes to reach the finish is relatively smaller?

Use the following sums for the computations: $\sum x_i = 3975$ min, $\sum x_i^2 = 159875$ min², $\sum (x_i - \bar{x})^3 = -628.12$ min³ y $\sum (x_i - \bar{x})^4 = 80701.95$ min⁴.