

Bachelor in Computer Science and Engineering

Statistics Problems

I Univariate Descriptive Statistics

1. The following table presents the cumulative frequency for the number of goals scored by the home team during the first 100 matches of the Spanish soccer league.

<i>Goles</i>	<i>Frecuencia absoluta acumulada</i>
0	30
1	
2	85
3	95
4	

If we know that the average number of goals of the home team during those 100 matches was 1,25:

- Add a column to the table containing all relative frequencies (no cumulative)
 - Find the median, the first and third quartiles, the interquartile range and draw the corresponding boxplot.
2. The weights in Kg corresponding to 10 students from the Carlos III University are:

59	63
70	83
70	83
86	84
90	69

Calculate without grouping the data:

- The average, the median, and the mode
- The quartiles
- The variance and the standard deviation
- Draw the boxplot
- Draw the histogram

II Bivariate Descriptive Statistics

1. The following statistical summary on a set of 50 data has been obtained for variables X and Y, with covariance 1,1 between them.

$$\begin{array}{llll} \sum x_i = 148 & \sum y_i = 219 & \sum x_i^2 = 480 & S_y^2 = 1.4448 \\ P_{25} = 2.19 & P_{50} = 3.03 & P_{75} = 3.64 & \end{array}$$

- The value $X=7$ could be considered an outlier?, and an extreme outlier? Justify your answer.
 - Obtain a lineal regression model which allows to estimate the value of Y by means of the values of X.
 - Estimate the value of Y when $X=5$ using the previous model.
2. An enterprise wants to know which is the level of employee satisfaction. They perform a satisfaction survey to ten employees whom are asked to assess from 0 to 10 the atmosphere in their workplace, being 0 an appalling atmosphere and 10 an unbeatable workplace climate. The number of years that the employees have been in the company has also been asked in the survey. It is assumed that the assessment depends on the number of years that the employee has been in the company and a regression line has been estimated:

$$\hat{y}_i = 5,8 - 0,22x_i$$

For each one of the next statements say if it is True or False

When the assessment is 0 the average time in the company is 5,8 years	
For each year in the company the assessment decreases by 0,2 points	
Just entering the company the average assessment is 0,2 points	
After a year in the company the assessment increases in 5,6 points in relation to the initial assessment	
After 10 years in the company the average assessment is 3,8 points	

3. When asking a set of 300 students we obtain that 10% are smokers and 40% female.
- Build the absolute joint frequency table (bivariate) for these 300 students considering variables Smoker/Non Smoker and Male/Female, knowing that the number of boys that smoke is equal to the number of girls that smoke.
 - Indicate the relative joint frequency distribution

- c. Indicate the relative marginal distribution for variable Smoker/Non Smokers
- d. Indicate the absolute frequency distribution for variable Smoker/Non Smoker conditioned on the individuals being girls.
- e. Which group smokes more, boys or girls?