TELECOMMUNICATIONS ENGINEERING

STATISTICS

2022-2023

ASSIGNMENT 1. DESCRIPTIVE STATISTICS

1. Analysis of a data set

The file `Popularity.mat` contains the information of a survey over 478 students from the US public school system about the factors that determine the popularity among the students. The survey collects information about the following variables:

1. **Gender:** 1=Boy, 2=Girl.
2. **Grade:** 4,5 or 6.
3. **Age:** Age in years.
4. **Race:** 1=white, 2=others.
5. **Area:** The school is in an 1=urban, 2=sub-urban, or 3=rural area.
6. **School:** School name.
7. **Goals:** Student's choice in the personal goals question where options were 1=Make Good Grades, 2=Be Popular, 3=Be Good in Sports.
8. **Grades:** Rank of "make good grades"(1=most important for popularity, 4=least important).
9. **Sports:** Rank of "being good at sports"(1=most important for popularity, 4=least important).
10. **Looks:** Rank of "being handsome or pretty"(1=most important for popularity, 4=least important).
11. **Money:** Rank of "having lots of money"(1=most important for popularity, 4=least important).

To analyse the file in MATLAB:

File → **Import data**, select `Popularity.mat`, → **Finish**

We have created in the `workspace` the matrix of size 478×11.

1. Calculate the frequency table of variable `Area`. The table must include the absolute, relative, cumulative absolute and cumulative relative frequencies. In which of the three types of areas most students are concentrated? **(1.5 points)**

2. What is the proportion of boys and girls? Represent graphically that proportion with a bar and a pie chart. What is the proportion of boys and girls whose schools are established in urban areas? **(1.5 points)**

3. Do histograms of the variables `Grade` and `Age` by the variable `Goals`. Calculate the mean and the standard deviation of the variables `Grades` and `Sports` by `Age` groups. **(1.5 points)**

4. Analyse the variables `Gender` and `Goals` in a double entry table. Calculate the absolute frequency table with its marginal distributions and the relative frequency table with its marginal distributions. **(1.5 points)**

2. Linear Transformations

1. Change of units. Consider the matrix `internet` in the file `internet.mat`, and consider the variable `MB` ("downloaded Mb"). Define a new variable, `KB`, as the nº of downloaded Kb, recall "1Mb = 1024Kb". The new variable is the result of a linear transformation of the form $y = a + bx$. From this transformation, check with MATLAB/Octave the next theoretical relations: **(2 points)**

   a) $\overline{y} = a + b\overline{x}$.

   b) $y_{\text{med}} = a + bx_{\text{med}}$, where med is the median.

   c) $s_y^2 = b^2 s_x^2$, where s^2 is the sample quasi-variance.

   d) $s_y = |b| s_x$, where s is the sample quasi-standard deviation.

2. Standardization of variables. Consider the variable `MB`, and denote it as x. Define a new variable y as the result of the standardization of x. The standardization consist to apply a linear transformation such that subtracts the mean value and divides by its standard deviation. The resulting variable has zero mean, and standard deviation and variance equal to one. **(1 point)**

   a) Determine the values of a and b of the corresponding linear transformation $y = a + bx$.

   b) Obtain the new standardized variable y and check in MATLAB/Octave, the next results: $\overline{y} = 0$, $s_y^2 = 1$ y $s_y = 1$.

3. Correlation between linearly transformed variables

Change of units. Consider the matrix `internet`, and variables `MB` (x ="downloaded Mb") and `connection` (v ="connection time in hours"). From them, create two new variables: y ="nº of downloaded KB" and u ="connection time in seconds". Note that you are applying a linear transformation of the type: $y = a + bx$ y $u = c + dv$, or simply a change of units: $a = c = 0$ and $b, d > 0$.

Check in MATLAB/Octave, the next result:

$$\rho_{y,u} = \frac{bd}{|b||d|} \cdot \rho_{x,v} = \frac{bd}{bd} \cdot \rho_{x,v} = \rho_{x,v}$$

which indicates that the correlation coefficient between two variables does not change if a change of units is applied. **(1 point)**