



## PRACTICE UNIT 5. DESCRIBING UNIVARIATE DATA WITH NUMERICAL MEASURES: DISPERSION AND SHAPE MEASURES

### OBJECTIVES

The objectives of this practice are:

- Handle data obtained in a survey.
- Handle big amounts of data and learn some basic statistical functions available in Excel.
- Calculation of dispersion and shape measures to describe data.
- Analysis (numerically and graphically) of the shape (skewness and kurtosis) of the distribution.
- To detect the existence of outliers using box-plots

### DATA

We are going to work with the data sets already employed in the practices of Units 3 and 4. The associated frequency distributions, graphic presentations and location measures calculated in these practices are included in the corresponding sheets.

### PART A. UNIVARIATE DESCRIPTIVE ANALYSIS FROM THE FREQUENCY DISTRIBUTION

**Exercise U5\_A1.** (Sheet 'A\_Place of residence')

- a) Using the frequency distribution provided for the variable *Residence* (Place of residence) perform a descriptive study, calculating the appropriate dispersion measures for this variable and interpreting them in the context of the given data.

**Exercise U5\_A2.** (Sheet 'A\_Number of siblings')

Using the frequency distribution provided for the variable *Siblings* (Number of siblings) answer to the following questions:

- a) What can be said about the variability of the distribution of the number of siblings of the students (included the student)?



- b) What is the change in the standard deviation of the variable that counts the number of brothers and sisters (not including the student)?

**Exercise U5\_A3.** (Sheet 'A\_Height')

Using the frequency distribution provided for the variable *Height* (Height of the student):

- a) What is the width of the interval of heights that contains the middle 80% of the female students? And that of male students?
- b) Is the average value representative for the whole group of students?
- c) What value is more extreme, a student 150 cm height or a student 190 cm height?
- d) Analyse the shape of the height distinguishing by gender using the graphic presentations obtained in Practice U3? Is this shape in accordance with the values of the central location measures calculated in Practice U4?

**Exercise U5\_A4.** (Sheet 'A\_Balance')

The following table shows the quarterly average balance (in euros) in the bank accounts of 75 clients of a bank branch:

Quarterly average balance (euros)	Number of clients
0-600	10
600-1200	15
1200-1800	35
1800-3000	10
3000-6000	5

- a) Is the average quarterly average balance representative of these 75 clients?
- b) If the bank has decided to charge a management fee of 20€ per quarter, what is the representativeness of the new average value?
- c) Analyse the shape of the distribution comparing the values of the average and the median.
- d) Analyse the shape of the distribution using the graphic presentation obtained in Practice U3.



## **PART B. UNIVARIATE DESCRIPTIVE ANALYSIS WHEN THE COMPLETE DATA SET IS AVAILABLE**

### **Exercise U5\_B1.** (Sheet 'B\_Number of siblings')

Using the values of the variable *Siblings* (Number of siblings) answer to the following questions:

- a) What can be said about the variability of the distribution of the number of siblings of the students (included the student)?
- c) What is the change in the standard deviation of the variable that counts the number of brothers and sisters (not including the student)?

### **Exercise U5\_B2.** (Sheet 'B\_Height')

Using the values of the variable *Height* (Height of the student):

- a) What is the width of the interval of heights that contains the middle 80% of the female students? And that of male students?
- b) In which group of students is height more homogeneous, in the male or in the female student's group? Are these average values representatives for the respective groups of students?
- c) Who is taller in relative terms, a male student 180 cm height or a female student 170 cm height?
- d) What value is more extreme, a male student 150 cm height or a female student 190 cm height?
- e) Analyse the shape (skewness and kurtosis) of the height distinguishing by gender. Calculate the appropriate measures.
- f) Use a box plot to present graphically the height of male and female students and comment on the result: Are there any outliers? Do the values of the variable follow a symmetric distribution or are they skewed?