Name and surname \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**A simple experiment**

A simple experiment was carried out among students enrolled in Medical Statistics course. It was asked to all the students to answer two questions: A. Read the following text, from a 2013 interview with Mia Couto, and count how many times the letter 'n' appears; B. write here your height in cm. Out of 134 students registered for the class, 106 answered and their answers are registered on the file: studentdata.xlsx

Check that data: how many variables do you have in the data?

 how many observations?

**Now focus on the first question: "A. Number of “n” in the text (from a 2013 interview with Mia Couto). Variable name:** No\_of\_n

1. Describe the data providing summary indicators.

Mean Standard deviation

Median Quartiles

Modal value

1. Are there any missing values? If so, how many are there?
2. Based on these data, give a point and an interval estimate of the number of “n” in the text
3. Display and describe the distribution of No\_of\_n.
4. Do you find any extreme/strange value? If so, which one?
5. The true value of “n” in the text is 70. Is the estimate of the number of “n” of point 3 biased?
6. Provide an estimate of precision and accuracy of the measure of the number of “n” in the text.

**Now focus on the second question: B) Height of students**

1. Describe the height (height\_cm) in the sample providing summary indicators.

Mean Standard deviation

Median Quartiles

Modal value

1. Are there any missing values? If so, how many are there?
2. Do you find any extreme/strange value? If so, evaluate whether it is a likely error or not.

In case, correct it creating a new variable corrected and call it “height\_adj”

Provide explanation of how did you change it:

1. Describe the height in the sample with the corrected data.

Mean Standard deviation

Median Quartiles

Modal values

Do they change with respect to those in point 1?

1. Describe the distribution of height at point 3 (corrected data).
2. Is the distribution approximately Gaussian?

Estimate an interval in which you would expect 95% of the observations based on the corrected data and compare it with the actual frequencies of subjects in the same interval.

1. Is the distribution bimodal? If so, provide a possible explanation.
2. Let’s take a random sample of size 15, report the following quantities on the course web-page:
	1. Mean $\overbar{x}=\sum\_{i=1}^{n}x\_{i}/n$
	2. Sample standard deviation $s=\sqrt{\frac{\sum\_{i=1}^{n}(x\_{i}-\overbar{x})^{2}}{n-1}}$
	3. $s\*=\sqrt{\frac{\sum\_{i=1}^{n}(x\_{i}-\overbar{x})^{2}}{n}}$