**Exercise 3, 2024. Statistics**

We have the attached database (<https://www.uv.es/mperea/readingskill.sav>) which contains data on reading ability through the PROLEC test in high school students, as well as the average number of words learned incidentally across several classes.

*1. Do you observe differences in reading ability (using the global measure of the test: PROTOTAL) between boys and girls?*

*2. Overall, does the PROLEC test (global measure PROTOTAL) follow approximately a normal distribution? What is the asymmetry? What about kurtosis?*

*3. If we want to select the best 15% of readers for a creative writing course, what cut-off point will we choose?*

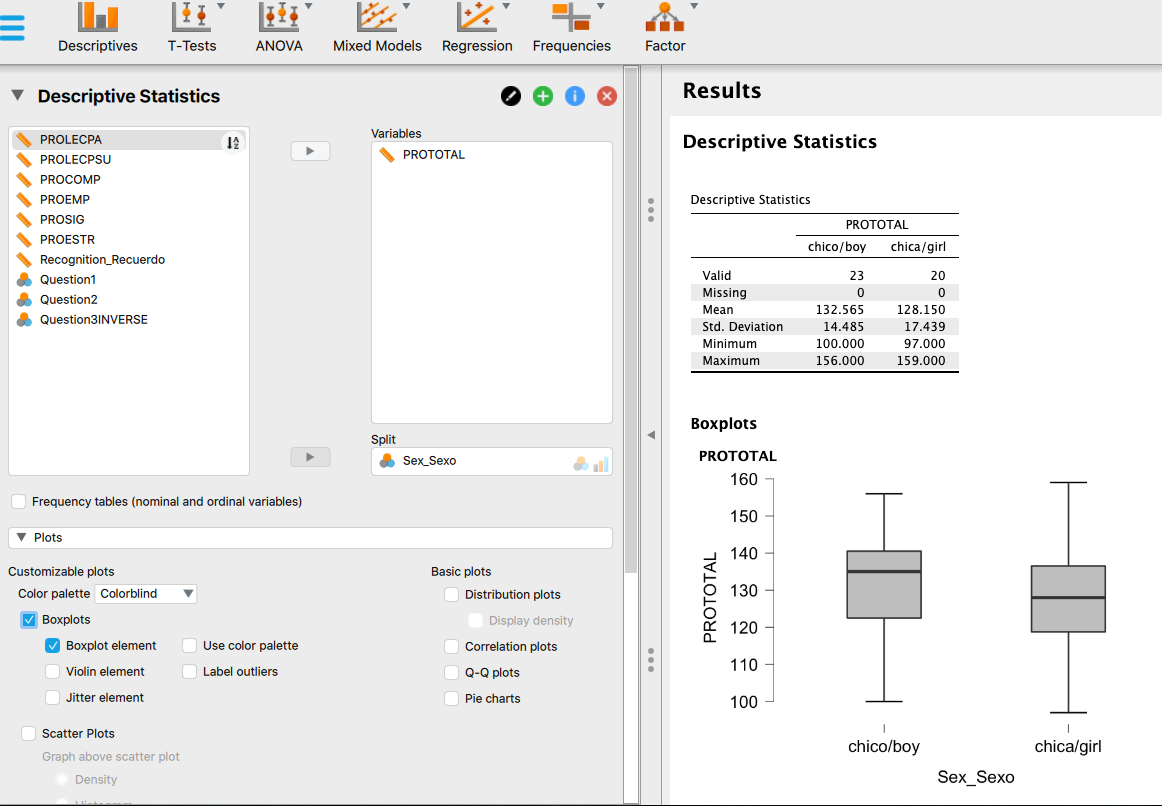
*4. If we want to select the worst 10% of readers for a reading ability improvement program, what cut-off point will we choose?*

*5. What is the standard score in PROTOTAL for the person in the first row in the file and what does it mean?*

**Answers**

*1. Do you observe differences in reading ability (using the global measure of the test: PROTOTAL) between boys and girls?*

The independent variable would be sex and the dependent variable would be reading ability in PROLEC (PROTOTAL).

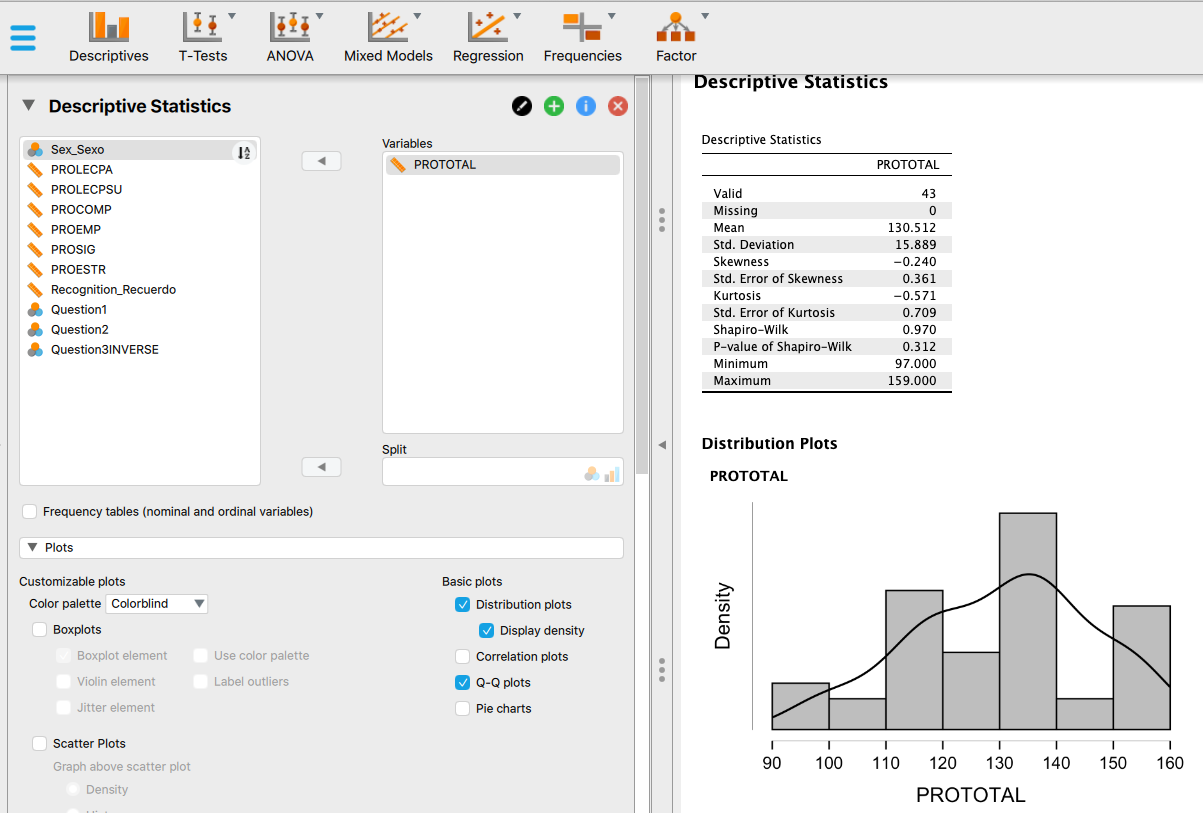


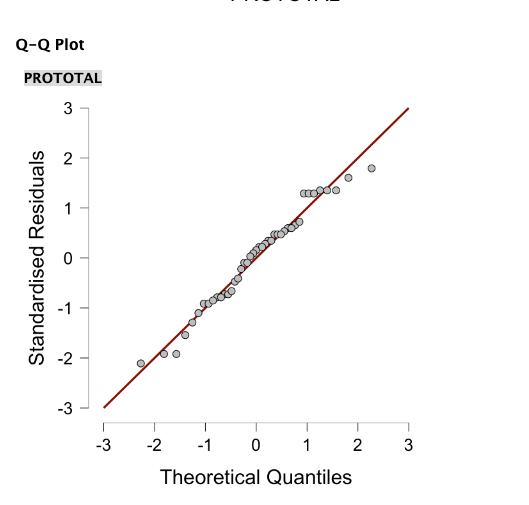
Boys had, on average, slightly higher reading ability scores than girls in this sample (M = 132.6 vs. 128.2, respectively).

(We used the mean; with the median the effects were similar, as shown in the box plot above.)

*2. Overall, does the PROLEC test (global measure PROTOTAL) follow approximately a normal distribution? What is the asymmetry? What about kurtosis?*

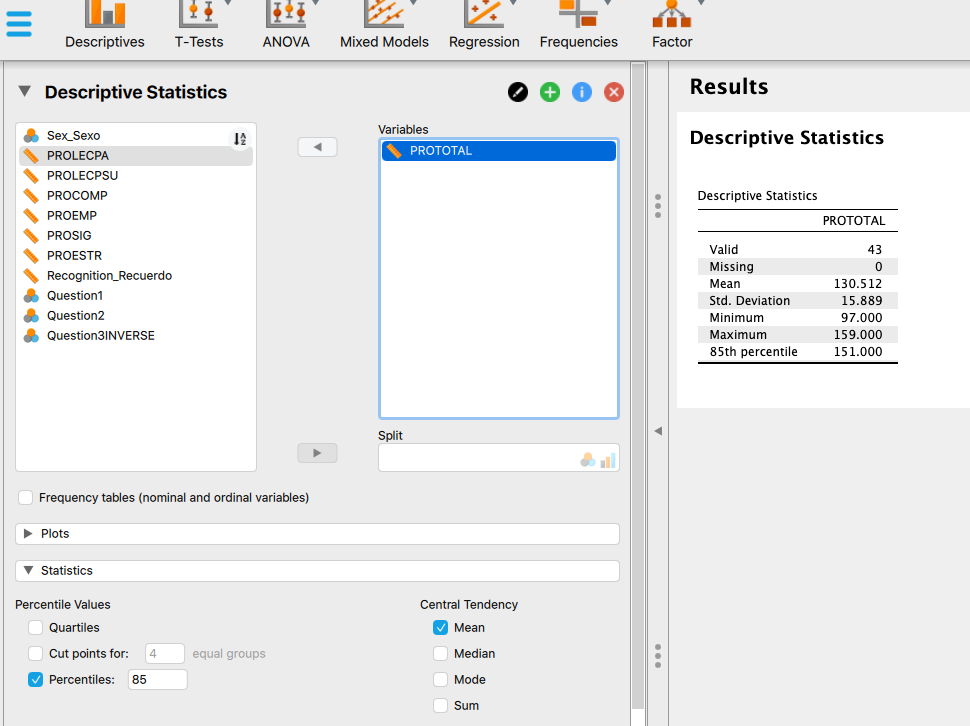
The most practical idea is to look at the histogram, in particular the smoothed line, in which a slight negative asymmetry can be seen. This is in fact is confirmed by the skewness index, which is -.240. Kurtosis is -.571 which suggests that the distribution is somewhat platykurtic.



However, the Q-Q plot shows that the points are not very far from the line, which suggests that the PROLEC distribution is not really very far from the normal distribution. Anyways, the fit is far from perfect… a larger sample would be required to have a definite answer.

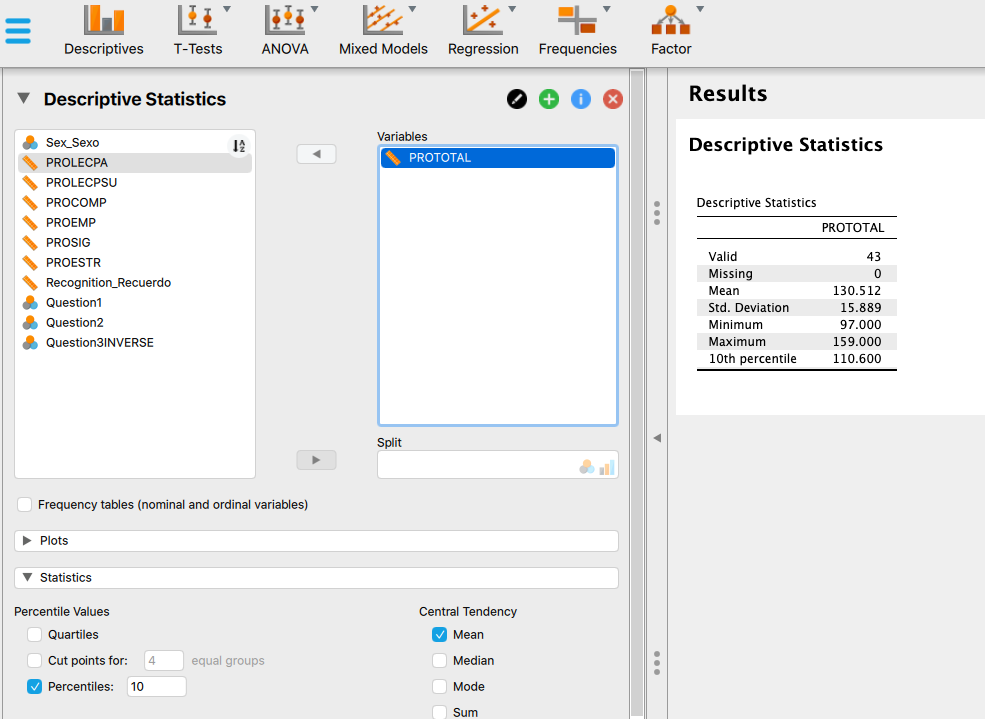
*3. If we want to select the best 15% of readers for a creative writing course, what cut-off point will we choose?*

We have to choose the 85th Percentile. This value leaves below 85% of the people and therefore, above this value, there will be 15% of the people. In our case, the cutoff is 151.



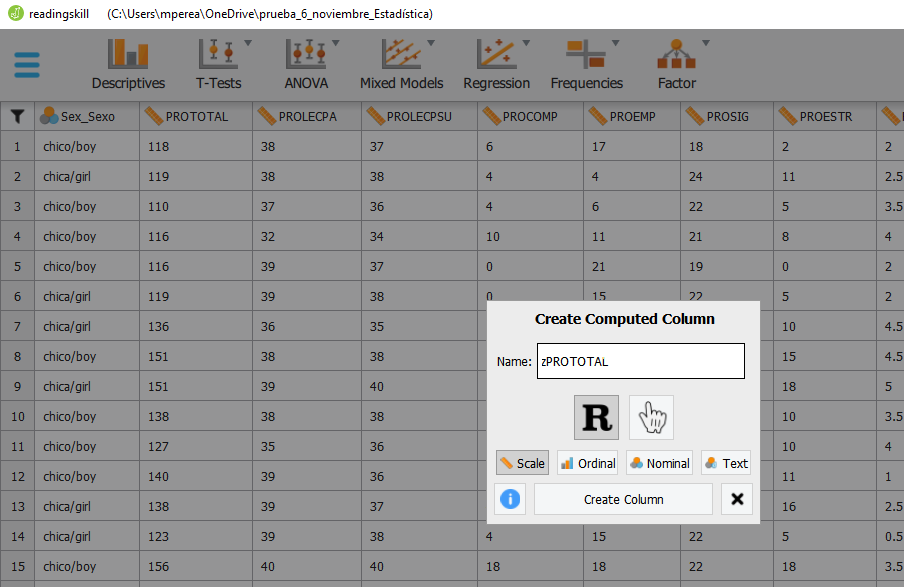
*4. If we want to select the worst 10% of readers for a reading ability improvement program, what cut-off point will we choose?*

We have to choose the 10th percentile. This value leaves below it 10% of the people. In our case it is 110.6.



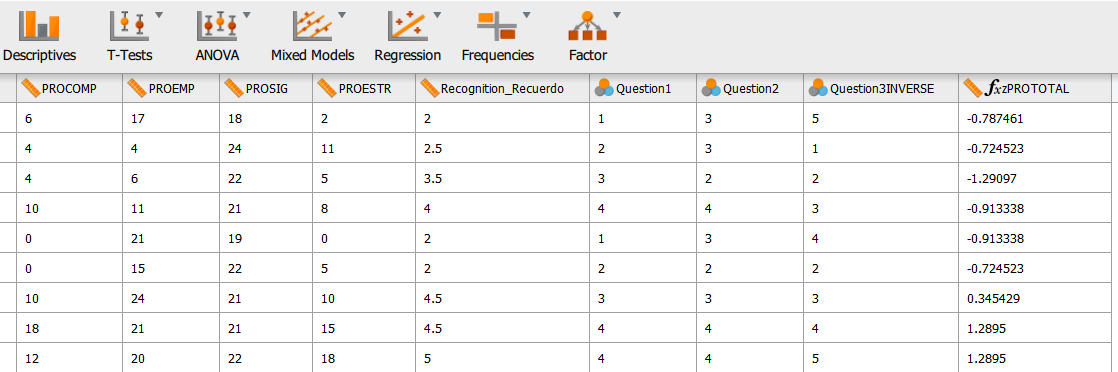
*5. What is the standard score (z-score) in PROTOTAL for the person in the first row in the file and what does it mean?*

To get typical scores, we have to follow the formula



The formula is:

scale(variable\_to\_be\_standarized)



The value of z for the first person is -0.78. This means that this person is 0.78 standard deviations below the average in reading ability. So this person has a reading ability somewhat lower than the average but not extremely low.