

The importance of descriptive analysis

A importância da análise descritiva

ÁLIDA ROSÁRIA SILVA FERREIRA¹ 

Statistical analysis is an important aspect of the scientific studies, and contributes to decision making. Considering this, it is relevant that the choice of statistical tools is of utmost importance, as well as the way the results are presented to the reader. A study to generate reliability needs to follow some steps: adequate sample size, correct choice of the participants and the appropriate method of analysis¹.

The choice of the adequate statistical method is based on two major premises: parametric and nonparametric data². The concept is that if the data distribution, assessed by an adequate statistical test, has a normal distribution, the mean and median have very close values, and both are representative of the groups to which they belong. When the sample turns out to be non-parametric (non-normal), the mean behaves differently from the median, and therefore it is no longer representative.

This definition is important because there are countless articles that, after defining that the data have a non-parametric distribution, and having selected the appropriate test, end up presenting a descriptive analysis of the variable using the mean and the standard deviation. The problem with this option is that the non-pa-

rametric tests do not use these values to generate their results, precisely because they contain such bias in the calculated values.

When combining the p-value generated by a non-parametric test and a descriptive data that uses the mean and the standard deviation, it is common for the reader to question the reason why a certain difference is not significantly in accordance to the test criteria. The reason is that the mean and standard deviation are highly sensitive to the smallest changes in the data.

As for the non-parametric data, it is important that the p-value is presented together with the median and the quartiles, and not with the mean and the standard deviation. This is not a problem when the data is parametric, therefore there is no harm in choosing any one of them.

The main point of the current discussion is to understand that a wrong choice in the presentation of the data can generate doubts as to the quality of the presented results, and may lead to conclusions that will be unreliable. A basic reliable descriptive statistic will be able to confirm what a correct and more sophisticated analysis tool has generated.

REFERENCES

1. Hair JF, Tatham RL, Anderson RE, Black W. Multivariate Data Analysis. 5th ed. New Jersey: Prentice Hall; 1998.
2. Vieira S. Bioestatística: tópicos avançados. Rio de Janeiro: Elsevier; 2003.

Received in: 18/06/2020

Accepted for publication: 19/06/2020

Conflict of interest: no.

Funding source: none.

Mailing address:

Álida Rosária Silva Ferreira

E-mail: alida.rsf@gmail.com



1 - Universidade Federal de Minas Gerais, Estatística/Demografia/Nutrição - Belo Horizonte - MG - Brasil