Statistical techniques and software used in the field of Clinical Medicine: a bibliographic review

Gerard Boxo[1], *Rosa Abellana*[1], *Josep L Carrasco*[1]

[1]jlcarrasco@ub.edu, Biostatistics, Department of Basic Clinical Practice, University of Barcelona

In the Health Sciences courses, students are typically introduced to Statistics through a basic core subject. The course syllabus varies according to the number of ECTS in each course. However, regardless of the teaching time, the syllabus includes statistical inference techniques, hypothesis contrasts testing and modeling approaches. Often, the statistical techniques presented in these subjects are heirs to teaching plans defined years ago.

Additionally, among the different teaching activities included in these subjects, we find the computer laboratories, in which students get in contact with statistical data analysis. This activity involves the use of statistical software that allows data management and the execution of appropriate statistical analyses. Currently, the range of statistical programs is very wide and the teacher must decide which software will be chosen. The factors for making this decision can be several. Examples are the cost, the teacher's familiarity with the software or the students' access to the software outside the classroom.

The objective of this work is to evaluate which are the statistical techniques, as well as the statistical software, that researchers in the field of Clinical Medicine are currently using. Thus, it will be possible to assess if the syllabus of Statistics basic subjects concurs with the methodology that the future healt professionals will meet in their research practice.

To achieve the objective proposed here, a bibliographic search will be carried out. Research papers published in 2022 will be downloaded by means of web scrapping using the Selenium Python's package. The journals considered are The Lancet, The New England Journal of Medicine, the Journal of the American Statistical Association and the British Medical Journal. These are the four top rated journals in the General Medicine category according to their impact factor. The information about statistical techniques, software and type of research design (experimental or observational) will be extracted from the downloaded papers using text mining techniques. Specifically, an algorithm of unsupervised autonomous learning will be applied.

Keywords: Text mining, Clinical Medicine, Statistical techniques, Statistical software