Multivariate Statistics

Assignment

Deadline: 9:00, 30-11-2022 (Wednesday, week 6), submit on Canvas in pdf format.
If the assignment is made in a group of two, then one submission is enough.
Submissions by (e-)mail will be accepted after the deadline,
in this case penalty of 3 (out of 10) points will be applied

General Instructions

- Write a report of at most 4 pages (12pt font, no double columns) in the form of a small article, i.e. brief introduction, data, methods, results (conclusion section is not needed) in which you answer a research question.
- Choose a topic from the options below
- Justify your statements/conclusions by reporting appropriate results (possibly in tables and/or figures).
- Do NOT copy and paste parts of other texts, write everything in your own words. NO PLAGIARISM.
- You can make a title page with: (in a large font) your name(s), your student number(s), and the title of the report. (This page is not counted for the number of pages)
- Add your code in Appendix (This is not counted for the number of pages)
- If you have questions, do not hesitate to ask them during the zoom meeting on Friday.

Option 1

We consider the dataset "DoctorVisits" from the `AER` package in `R`. Alternatively, it can be found in the Journal of Applied Econometrics data archive `http://qed.econ.queensu.ca/jae/1997-v12.3/mullahy/`. Check the documentation. The goal is to reanalize these data using the methods from the course. Find an appropriate model and estimation method. Formulate a research question that you find important/interesting.

Option 2

What is the effect of outliers on the performance of Hotelling's T^2 statistic? Design a simulation study and investigate the influence of outliers on the level of the test and on the power of the test. Can Minimum Covariance Determinant estimator solve the problem?

Option 3

In economics and applied econometrics it is common to make various robustness checks and/or specification tests. One of the problems to be checked is heteroskedasticity in regression. There are various tests, one well-known example is Breusch-Pagan test. The goal is to investigate what is the effect of outliers on the performance of the test(s) of heteroskedasticity. What happens with the level of the test, power of the test, p-values? Design a simulation study to investigate this issue.