Data Analytics: Statistical Programming
Exam

Michael Lechner Arnau Valladares-Esteban

Friday 20th December, 2019

BEFORE starting the exam

- Take out your student ID card and place it on the table.
- Make sure your phone is turned off.
- Make sure your machine is connected to a power source or that you have enough battery to complete the exam.
- Make sure your connection to the Internet works.
- Make sure you can log in to the Canvas site of the course.

Exam instructions

- You have ninety minutes to complete the exam.
- Submit the answers in Python as one Jupyter Notebook.
- Submit the answers in Gauss in two files: one containing code and another containing the output (either ASCII files or copy/pasted in a text processor document).
- Provide a nice output for your answers.
- You can look up online documentation (language and libraries).
- You **cannot** look up any code used in class, code created by you in the past, tutorial, forum or similar.

In Barcelona, high-school students need to pass an exam before entering University. The exam is graded on a scale from 0 to 10. Grades are used to assign students to the different degrees offered by universities (Economics, Business Administration, Medicine, Engineering, etc). The student with the highest grade is the first to be assigned to the degree of their choice. When a degree has all its spots assigned, the next student in the grade ranking is assigned to their second choice. This procedure continues until all students with a grade above 5 are assigned to a degree.

Given the crucial importance of this exam, a group of researchers has gathered data to investigate the determinants of students' grades. The dataset `DadesPAU.cvs/.dat` contains the following information about a sample of 9,738 students from Barcelona:

- `IQ` is a measure of the cognitive abilities of each student.
- `HStudied` contains the cumulative hours spent studying by each student the 12 months prior to the exam.
- `QualityHS` is a measure of the quality of the high school attended by each student.
- `Grade` final grade in the exam.

Questions

Answer all questions both in Gauss and in Python:

1. Load the data and describe all variables. In particular, show their means, standard deviations, and medians.
2. Compare the medians computed by the built-in functions with your own implementation of a function that computes the median. A naive algorithm (which uses a built-in sorting function) is enough, do not worry about computing time.
3. Investigate the joint relationship of `IQ`, `HStudied`, and `QualityHS` on the grade achieved in the exam (`Grade`) by running an OLS regression (it's sufficient to use the point estimates). Discuss the results.
4. Develop your own code for the point estimator of OLS. Compare the estimates of your code with those by the built-in functions.