Data Analytics: Statistical Programming
EXAM - HS 2020

Michael Lechner Arnau Valladares-Esteban

18 December 2020

Before you start 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 is stable.
- ▷ Make sure you can log in to the Canvas site
- ▷ Carefully read the exam instructions from below.
- ▷ Wait until the exam supervisor starts the exam.

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 file or copy/pasted in a text processor document).
- ▷ Provide a nice output for your answers. We can only grade comprehensible results.
- ▷ You **can** look up official documentation of language and libraries (online or within Gauss and Python).
- ▷ You **cannot** look up any code used in class, tutorials, code created by you in the past, nor any web forum or similar.
- ▷ Submit your answers in the respective module on Canvas. Submit all related files to your solutions. You have only one try to submit your files.

Information on the data

In a famous article, Card and Kruger (AER, 2014) analyse the impact of a an increase in the minimum wage on employment. The authors find that an increase in the minimum wage set leads to an increase in the amount of workers employed, a result that contradict economic theory.

Card and Krueger study two neighbouring states in the U.S., New Jersey and Pennsylvania. In 1992, state legislators in New Jersey increased the minimum wage while Pennsylvania did not. Card and Krueger surveyed fast-food restaurant on the two sides of the border between New Jersey and Pennsylvania, before and after the policy change.

The dataset *CardKruger.csv* contains the sample of 716 observations with the following variables:

- ▷ *State*: 1 for New Jersey, 0 for Pennsylvania.
- ▷ *Year*: Date of the data collection, either 1992 or 1993.
- ▷ *Wage*: Lowest paid wage to any employee.
- ▷ *FTE*: Number of full-time employees.
- ▷ *PTE*: Number of part-time employees.

Important remark: The provided dataset has been slightly modified with respect to the original to fit the exam questions. Your results do no have to replicate the original estimates.

Exam questions

Answer the following questions in Gauss **and** Python. The stated points per exercise can be attained in each programming language individually. The maximal possible amount of points is therefore 24 Points.

Ex.1 [2 Points] Load the data and describe every variable by min, mean and max.

Ex.2 [2 Points] Verify that the minimum wage in New Jersey has risen from 1992 to 1993.

Ex.3 [3 Points] Calculate the standard deviation of the reported wages individually for both states with your own implemented function which does not rely on an in-built function that determines the standard deviation.

Help: The standard deviation is a measure of dispersion of a set of values. It is given by $\sigma(x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$*, where* $\bar{x}$ *is the average of* x*.*

Ex.4 [3 Points] Estimate the effect of the increased minimum wage on the number of full-time employees with the Difference-in-Difference estimator. Discuss your results.

Help: The Difference-in-Difference estimator (DiD) is a commonly used estimator. It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group. In our case: $\tau_{DiD} = (\bar{x}_{NJ,1993} - \bar{x}_{NJ,1992}) - (\bar{x}_{PEN,1993} - \bar{x}_{PEN,1992})$*, where* $\bar{x}_{u,v}$ *is the average outcome in state* u *for a given year* v*.*

Ex.5 [2 Points] The Difference-in-Difference estimator is in this particular case equivalent to the linear model $FTE = \alpha + \beta \cdot Year + \gamma \cdot State + \tau_{\text{OLS}}\, Year \cdot State$. Estimate this model with the OLS estimator (you can use in-build functions) and compare the results to what you obtained in **Ex.4**.