Problem Set 4

SOC-GA 2332 Intro to Stats (Fall 2023)

Due: Monday, Dec. 18th, 11:59 pm

Instructions

1. Submit two files for each problem set. The first is a **R Markdown** (`.Rmd`) file that can be run without error from start to end. The second is a **PDF** rendered from your R Markdown file or created using LaTeX.

2. Name your files following this convention: `[Last Name]_ps1.Rmd` and `[Last Name]_ps1.pdf`.

3. Both files should be submitted to the TA via e-mail (wj2068@nyu.edu) before the time specified above.

4. Comment on your code wherever possible and explain your ideas in detail. You will get credit for showing the steps you take and for explaining your reasoning, even if you do not get the correct final result.

Part 1: Fixed Effects and Random Effects Models

In this exercise, we will use the dataset `sibling_data.dta` to study the effect of mother experiencing stress in pregnancy on a child's birth weight. All variables are described in the table below.

Variable Name	Variable Detail
Group Variable	
`householdid`	Unique household id; Siblings from the same household share the same household id
Dependent Variable	
`birthwt`	birth weight measured in pounds
Independent Variable	
`stress`	Whether or not mother experienced stress in pregnancy (Yes=1; No=0)
`female`	Infant's gender (Female=1; Male=0)
`magebirth`	Mother's Age at Birth
`numsibling`	Number of Sibling
`meduy`	Mother's Years of Schooling
`feduy`	Father's Years of Schooling

1.[20pts] Import the dataset `sibling_data.dta` to your R environment. Including all independent variables, build an OLS model, a fixed effects model, and a random effects model (you can use the `plm` R package). Summarize regression results in a table. Make sure to label your models in your table (the `column.labels` argument in `stargazer()`).

Hint: `householdid` *is used as a grouping variable in your FE or RE model.*

2.[5pts] Interpret the coefficient of `stress` in the OLS model.

3.[5pts] Interpret the coefficient of `stress` in the fixed effects model.

4.[5pts] Interpret the coefficient of `magebirth` in the fixed effects model.

5.[5pts] Perform a F-test to compare the OLS and the fixed effects models and interpret outputs (use `pFtest` from the `plm` R package).

Hint: You may check this website *for the F-test in comparing OLS and FE models*

6.[5pts] Compared to an OLS model, what are the benefits of using a fixed effects model?

7.[5pts] Why don't we get coefficients of `meduy`, `feduy`, and `numsibling` in the fixed effects model?

8.[10pts] Why do we get coefficients of `meduy`, `feduy`, and `numsibling` in the random effects (random intercepts) model?

9.[10pts] How does the coefficient of `stress` change between (a) the OLS model and the fixed effects model and (b) the OLS model and the random effects model? What could be the potential causes that lead to these changes?

10.[10pts] Perform a Hausman test to compare fixed-effects and random-effects models (figure out how to use the `phtest` function from the `plm` R package; you might look into the help page and examples by typing ?plm::phtest into your console)

  (a) What is the null hypothesis of this model? What is the alternative hypothesis?

  (b) Is the null hypothesis rejected? Based on the test, which model would you use?

Hint: Hausman test evaluates whether the results (i.e. the estimated coefficients) from a fixed-effects and random-effects model are significantly different. You are encouraged to search online what the test is and how to decide which model is preferred.

Part 3 Data Transformation

[20pts] Import `parent_inc.csv` to your R environment. The data frame looks like this:

famid	father_name	mother_name	father_income	mother_income
1	Arthur	Jess	42000	45000
2	Harry	Pam	35000	24000
3	Matt	Mary	78000	55000

Then, use either `tidyverse` functions with the piping syntax (*i.e.*, `pivot_longer()` or `pivot_wider()`) or the `data.table` functions (*i.e.*, `dcast` and `melt`) to transform the data frame to the following structure:

famid	type	name	income
1	father	Arthur	42000
1	mother	Jess	45000
2	father	Harry	35000
2	mother	Pam	24000
3	father	Matt	78000
3	mother	Mary	55000

Make sure to document the steps you take in your code and display the tidied data frame in your PDF document.

Hints: You may check Lab 1, Part 5 for details.