Weiyi Li

2020/05/14

Casual Inference

Homework 2

**Git and Github**

1.

“Git is a distributed version control system.” “GitHub is an online hosting platform that provides an array of services build on top of the Git system” with “similar platforms include Bitbucket and Gitlab”. Git could work without installing GitHub.

2.

The benefit of using git to organize is that by using the pull, push and command, it’s easier for us to do version control and leave some notes on different versions of the projects. It might be challenging for me because git is a new technique for me, and I might sometimes forget to do my script under git or forget to pull and push. I might start to try using it since now to get used to it to minimize the challenges.

3.

The 4 main commands are: Git shell, merge conflicts, branches and forking.

Git shell is mainly used to stage all files, updates, new files; commit changes; pull and push the repository. Merge conflicts is to merge with different versions of the files. Branch could be create new branch; push or list branches; switch branches; or delete branches. Forking means to clone a repo and branching from it.

4&5

<https://github.com/weiyili96/Titanic.git>

**OLS**

7f

After the decomposition, I think the true effect of first class on surviving the sinking is almost the true effect but might be a bit larger, but definitely much better than the non-clustering ones.

7g

When considering about the heteroskedasticity issue, whether sits in the first class and whether is female are positively correlated with survival; age is negatively correlated with survival. When not considering about the heteroskedasticity, we got the same result with

When considering about the heteroskedasticity, but with different confident intervals. This is because for the heteroskedasticity regression, the formula for estimating the standard errors changes when allowing for arbitrary serial correlation within group.

7h

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |  |
| VARIABLES | regression | Regression with heteroskedasticity | Regression with sex clustering | Regression with vce | Means of variables |
|  |  |  |  |  |  |
| first\_class | 0.232\*\*\* | 0.232\*\*\* | 0.354 | 0.354\*\*\* | 0.148 |
|  | (0.0254) | (0.0259) | (0.188) | (0.0288) |  |
| age | -0.136\*\*\* | -0.136\*\*\* |  |  | 0.950 |
|  | (0.0405) | (0.0497) |  |  |  |
| gender | 0.464\*\*\* | 0.464\*\*\* |  |  | 0.214 |
|  | (0.0220) | (0.0242) |  |  |  |
| Constant | 0.318\*\*\* | 0.318\*\*\* | 0.271 | 0.271\*\*\* |  |
|  | (0.0401) | (0.0493) | (0.123) | (0.0103) |  |
| Survival  (depended) |  |  |  |  | 0.0323 |
| Observations | 2,201 | 2,201 | 2,201 | 2,201 |  |
| R-squared | 0.239 | 0.239 | 0.072 | 0.072 |  |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

7i

When considering the heteroskedasticity issue, if the passenger is in first class, he will have more possibility to survive; if he’s not in first class, this passenger could also survive, but in a lower probability compared to the first-class passenger.

When clustering for the sex effect, we got the same result with the vce regression.

Holding age and gender effect constant, first-class passengers will have more possibility to survive; if this passenger is not in first class, he could also survive, even with a higher probability. And when considering about the heteroskedasticity effect, the results seem to be the same.

My answer to part f seems correct, because the effect of being a first-class passenger has decreased, which means that should be larger than the true effect as I said in f.

8

The 2 coefficients are the same, with a little difference on their standard errors. I think the coefficients don't vary a lot is because age and gender are not highly correlated with each other; therefore, those shouldn't be the disturbance when doing the long regression.