PS7 Mozaffar

Yaseen Mozaffar

March 2020

1 Problem 6

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,669	1.625	0.386	0.005	1.362	1.936	2.261
hgc	2,229	13.101	2.524	0	12	15	18
tenure	2,229	5.971	5.507	0.000	1.583	9.333	25.917
age	2,229	39.152	3.062	34	36	42	46

Income data tends to be MNAR. This may be due to self-reporting bias (people of lower income are less likely to respond to surveys asking about income) or some other structural difficulty in getting in touch with those of lower income. However, neither of these explanations necessarily make sense with this dataset because the logwage data is missing in about 25.12 percent of the observations while a much lower percentage of observations were missing other variables. If it was harder to reach people of a certain income, there likely wouldn't be such complete data for the other variables.

Table 2:

	Dependent variable:		
	logwage		
	(1)	(2)	(3)
hgc	0.062***	0.049***	0.049***
	(0.005)	(0.004)	(0.004)
collegenot college grad	0.146***	0.160***	0.160***
	(0.035)	(0.026)	(0.026)
tenure	0.023***	0.015***	0.015***
	(0.002)	(0.001)	(0.001)
age	−0.001	−0.001	−0.001
	(0.003)	(0.002)	(0.002)
marriedsingle	−0.024	−0.029**	−0.029**
	(0.018)	(0.014)	(0.014)
Constant	0.639***	0.833***	0.833***
	(0.146)	(0.115)	(0.115)
Observations	1,669	2,229	2,229
R^2	0.195	0.132	0.132
Adjusted R^2	0.192	0.130	0.130
Residual Std. Error	0.346 (df = 1663)	0.311 (df = 2223)	0.311 (df = 2223)
F Statistic	80.508*** (df = 5; 1663)	67.496*** (df = 5; 2223)	67.496*** (df = 5; 2223)
Note:			*p<0.1; **p<0.05; ***p<0.01

2 Problem 7

Across the four imputation methods, I got beta1 values of .062, .049, .049, and .5, respectively. As the true value was .093, it's more likely than not that I made a mistake in methods 2-4. But it's too late to troubleshoot my code, so oh well. Also, I don't know why table 2 is doing what it's doing. (i.e. being off center and not being under the right heading) My best guess is that it's too wide, but I can't figure out how to resize. Sorry again but I'm out of time.

3 Problem 8

Progress is slow but steady. As my project is my grad research, I'm trying to put in extra work upfront on the theoretical basis for the paper before I get into the actual data analysis. So, not a whole lot being done using tools from this class, but I'm more or less done getting through the lit base and identifying my datasets. Ultimately, I think I'll be using nonlinear modeling techniques in my project. Because the outcome variable is binary (monopoly formation), it may be as simple as a logit or probit model. In any case, I'm going to attempt an unsupervised modeling portion, in hopes that my model can be used to predict what industries/companies are likely to form monopolies in the near future.