

PS7

Conrad Polkosnik

March 13, 2018

1 Part 6. Stargazer Summary Table

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
logwage	1,669	1.625	0.386	0.005	2.261
hgc	1,669	12.556	2.322	0	18
tenure	1,669	5.225	5.095	0.000	24.750
age	1,669	39.171	3.085	34	45

1.1 Analysis

I would guess that the observations are MCAR, as it just so happens that those values aren't there from the sample of data on wages. There appears to be no relationship between the data points of which are missing from the sample and any of the values in the data set itself. The missing data seems to be just a random subset of data in this data set and is not conditional on any other variable.

2 Stargazer Regression Table for Both Models

Table 2:

	<i>Dependent variable:</i>	
	logwage	
	(1)	(2)
hgc	0.062*** (0.005)	0.062*** (0.005)
collegenot college grad	0.146*** (0.035)	0.146*** (0.035)
tenure	0.023*** (0.002)	0.023*** (0.002)
age	-0.001 (0.003)	-0.001 (0.003)
marriedsingle	-0.024 (0.018)	-0.024 (0.018)
Constant	0.639*** (0.146)	0.639*** (0.146)
Observations	1,669	1,669
R ²	0.195	0.195
Adjusted R ²	0.192	0.192
Residual Std. Error (df = 1663)	0.346	0.346
F Statistic (df = 5; 1663)	80.508***	80.508***
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

2.1 Analysis

I am not 100 percent sure if I created the models correctly, but I produced the same statistics and beta values for both models, according to the stargazer regression table. Although I used different commands and different packages to produce the regression models, I still received the same values. For the value of B1, I got 0.062. This value underestimated the coefficient for hgc as the true value was equal to 0.093. In regard to what I believe the mean imputation method has done, I believe it was a major player to underestimating the coefficient, B1. In my opinion, I believe it would be better to underestimate data than to overestimate as if the actual data that was produced in the future was better than what one had hoped for, it just becomes somewhat of a bonus to the creator of the model, rather than a let-down. Underestimating provides somewhat of a reserved and more likely analysis.

3 Project Update

For my project, as you may already know from the previous homework assignment, I am going to forecast the stock price for Exxon Mobil. I plan to simulate some number of trading days using, what is known as, the "Random Walk" method and then using Monte Carlo simulation to produce some number of iterations to build confidence intervals for my forecast. The data I am pulling my historical prices for Exxon Mobil is from Yahoo, using the Quantmod package in R. I am a big fan of visualization and enjoy analyzing charts, so I plan to base my model around crunching the numbers and producing visuals so that anyone who is interested can completely understand exactly how and why I arrived at my forecasted price. After researching about some different modeling approaches, if I understand this correctly, I will be taking an agent-based modeling approach. This approach is defined as a class of computation models for simulating the actions and interactions of autonomous agents with a view to assessing their effects on a system as a whole.