

PS7

Jordan Hoehne

March 2018

Table 1: Question 6: Stargazer for removed variables

Statistic	N	Mean	St. Dev.	Min	Max
logwage	1,669	1.620	0.386	0.005	2.260
hgc	1,669	12.600	2.320	0	18
tenure	1,669	5.220	5.090	0.000	24.800
age	1,669	39.200	3.080	34	45

1 Question 7

The true value of $B^1 = 0.093$. Comment on the differences of B^1 across the models. What patterns do you see? What can you conclude about the veracity of the various imputation methods? Also discuss what the estimates of B^1 are for the last two methods.

2 Beta is off

The true value of $\beta = 0.093$. Comment on the differences of B^1 across the models. What patterns do you see? What can you conclude about the veracity of the various imputation methods? Also discuss what the estimates of B^1 are for the last two methods.

1. estimate the regression using only complete cases (i.e. do listwise deletion on the log wage variable ... this assumes log wages are Missing Completely At Random)

Omitting the missing observations gives us

2. perform mean imputation to fill in missing log wages

3. impute missing log wages as their predicted values from the complete cases regression above (i.e. this would be consistent with the “Missing at Random” assumption)

Filling in for the mean from the full data, the variance is reduced and we expect a

4. use the mice package to perform a multiple imputation regression model

3 Question 8: Project Update

At first, I thought about working with the iris data set to go over some of the fundamentals of data science. However, after working with it and doing some research, I didn't think it would be challenging enough or give me the "wow, that's amazing" impression in interviews. This led me to seek out another project that I've been wanting to work on lately... computer vision! I had been interested in writing a script that could read pdf's and transpose handwritten and printed words into a database after scanning the image, which would completely disrupt the data entry industry. Although faulty and no where near perfect, this field of study excites me. If you have seen Silicon Valley, then you know about the "hot-dog app." Similar to the hot dog app, I would like to train a neural net to classify images, perhaps on star wars data, Jedi vs. Sith or something similar like cat vs. dog or even easier, hot-dog or not hot-dog. I'll be using Keras on top of Tensorflow and have a model up and running by next Monday! I'm excited to start down this path and to get started with Deep Learning.

Table 2: complete cases

	<i>Dependent variable:</i>
	logwage
hgc	0.062*** (0.005)
collegenot college grad	0.146*** (0.035)
tenure	0.023*** (0.002)
age	−0.001 (0.003)
marriedsingle	−0.024 (0.018)
Constant	0.639*** (0.146)
Observations	1,669
R ²	0.195
Adjusted R ²	0.192
Residual Std. Error	0.346 (df = 1663)
F Statistic	80.500*** (df = 5; 1663)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3: Replacing Missing logwage with mean(logwage)

Statistic	N	Mean	St. Dev.	Min	Max
logwage	1,669	1.620	0.386	0.005	2.260
hgc	1,669	12.600	2.320	0	18
tenure	1,669	5.220	5.090	0.000	24.800
age	1,669	39.200	3.080	34	45