

PS7 *Gillingham*

Andrew Gillingham

March 2019

1

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

It would appear that there is almost 75 percent the count of logwage than the rest of the columns. So there would be about 1/4 missing rate.

2

The results I got for my β_1 in the first two tests did differ kind of a lot from the actual .093 that it was supposed to be. As for the β_1 of the final regression imputation with mice, it was significantly closer at .072. It would appear that the better models for imputing data will result in more accurate β values.

3

I'd like to think that my project is at least starting to come together. As we discussed the other week, I am probably going to be double dipping a little between this project and my thesis so I already have my data and programs that I can use that will also be able to do some analysis. I'm not sure about the full range of functions pybaseball and baseballr are able to perform yet, so over break I will be looking at those in order to figure out what questions I should ask and try and answer. However, I do feel at least a little more confident about this project than I usually do with it being a little more than a month away from being due.

Table 2:

	<i>Dependent variable:</i>		
		logwage	
	(1)	(2)	(3)
hgc	0.062*** (0.005)	0.050*** (0.004)	0.062*** (0.005)
collegenot college grad	0.145*** (0.034)	0.168*** (0.026)	0.145*** (0.034)
tenure	0.050*** (0.005)	0.038*** (0.004)	0.050*** (0.005)
I(tenure^2)	-0.002*** (0.0003)	-0.001*** (0.0002)	-0.002*** (0.0003)
age	0.0004 (0.003)	0.0002 (0.002)	0.0004 (0.003)
marriedsingle	-0.022 (0.018)	-0.027** (0.014)	-0.022 (0.018)
Constant	0.534*** (0.146)	0.708*** (0.116)	0.534*** (0.146)
Observations	1,669	2,229	1,669
R ²	0.208	0.147	0.208
Adjusted R ²	0.206	0.145	0.206
Residual Std. Error	0.344 (df = 1662)	0.308 (df = 2222)	0.344 (df = 1662)
F Statistic	72.917*** (df = 6; 1662)	63.973*** (df = 6; 2222)	72.917*** (df = 6; 1662)

Note:

*p<0.1; **p<0.05; ***p<0.01