

PS7

Alli Penner

March 2018

1 Summary Table

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
logwage	1,669	1.625	0.386	0.005	2.261
hgc	2,229	13.101	2.524	0	18
tenure	2,229	5.971	5.507	0.000	25.917
age	2,229	39.152	3.062	34	46

2

The log wage data is missing at a rate of 25.12 percent. It appears to be missing at random. There appears to be more missingness for women who are married, have high tenure, or college degree. None of this missingness is complete, so the data can reasonably be imputed.

Every method underestimated the true value of the parameter. Linear, PMM, and complete cases all got equally close however.

3

I have been really swamped so I haven't looked to closely at my project. But I want to work with that NFL combine data I had in the last homework. I'm not sure how I want to model it yet, but expect the interaction between predictors will be important. So Trees or a neural networks might be good routes.

Table 2:

	<i>Dependent variable:</i>		
	Complete (1)	logwage Mean (2)	Linear (3)
hgc	0.062*** (0.005)	0.050*** (0.004)	0.062*** (0.004)
collegenot college grad	0.145*** (0.034)	0.168*** (0.026)	0.145*** (0.025)
tenure	0.050*** (0.005)	0.038*** (0.004)	0.050*** (0.004)
tenure2	−0.002*** (0.0003)	−0.001*** (0.0002)	−0.002*** (0.0002)
age	0.0004 (0.003)	0.0002 (0.002)	0.0004 (0.002)
marriedsingle	−0.022 (0.018)	−0.027** (0.014)	−0.022* (0.013)
Constant	0.534*** (0.146)	0.708*** (0.116)	0.534*** (0.112)
Observations	1,669	2,229	2,229
R ²	0.208	0.147	0.277
Adjusted R ²	0.206	0.145	0.275
Residual Std. Error	0.344 (df = 1662)	0.308 (df = 2222)	0.297 (df = 2222)
F Statistic	72.917*** (df = 6; 1662)	63.973*** (df = 6; 2222)	141.686*** (df = 6; 2222)

Note:

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