

# PS7 McGuire

Kevin McGuire

March 10, 2020

## 1 Summary Results

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

Roughly 1/3 of observations are missing logwage. At this point, without further looking at the data, I would guess logwage is either MAR or MNAR.

## 2 Regression Results

Table 2:

	<i>Dependent variable:</i>		
	logwage		
	(1)	(2)	(3)
hgc	0.062*** (0.005)	0.049*** (0.004)	0.052*** (0.005)
collegenot college grad	0.146*** (0.035)	0.160*** (0.026)	0.167*** (0.027)
tenure	0.023*** (0.002)	0.015*** (0.001)	0.016*** (0.001)
age	-0.001 (0.003)	-0.001 (0.002)	-0.003 (0.002)
marriedsingle	-0.024 (0.018)	-0.029** (0.014)	-0.022 (0.014)
Constant	0.639*** (0.146)	0.833*** (0.115)	0.839*** (0.118)
Observations	1,669	2,229	2,229
R <sup>2</sup>	0.195	0.132	0.146
Adjusted R <sup>2</sup>	0.192	0.130	0.144
Residual Std. Error	0.346 (df = 1663)	0.311 (df = 2223)	0.320 (df = 2223)
F Statistic	80.508*** (df = 5; 1663)	67.496*** (df = 5; 2223)	75.982*** (df = 5; 2223)

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Over all of the regression models, the B1 value stays relatively similar. The value decreases across the subsequent non-mice models, though the range stays more similar within itself than the true value as presented in the assignment.

### **3 Final Project Update**

For my project, I have set up all of the data collection I plan to use. The data is being regularly collected from Twitter and FiveThirtyEight. I have yet to finalize any plans for modeling, but I imagine machine learning will be a key component.