

PROBLEM SET № 7

Grant Naberhaus, University of Oklahoma

03/10/2020

ProblemSet Questions

The logwage variable is present in approximately 75 percent of the observations after missing observations for hgc and tenure were dropped. The logwage variable is most likely MAR. All the models have somewhat similar, statistically significant, Beta1 values. These values are much different from the true value, which seems odd. The last two models appear to have similar beta values, but not identical.

Project Progress

I'm still interested in constructing an NBA model. My initial expectations of creating a predictive model may have been a little too ambitious given my limited machine learning knowledge. I recently read about a model that used the sum of player efficiency ratings for a team (minutes adjusted), and correlated this sum with win percentage. Using this as a basis, I would also add in days rest as binary variables and include home court as an additional binary variable. Additionally, I considered trying to include an elevation variable, but I figured this would exhibit a high multicollinearity with the home court variable, and only two NBA teams play at a significantly higher elevation (Denver Nuggets and Utah Jazz).

Tables

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

Table 2: Results

	<i>Dependent variable:</i>		
	logwage		
	(1)	(2)	(3)
hgc	0.062*** (0.005)	0.050*** (0.004)	0.060*** (0.004)
collegenot college grad	0.145*** (0.034)	0.168*** (0.026)	0.134*** (0.025)
tenure	0.050*** (0.005)	0.038*** (0.004)	0.047*** (0.004)
tenuresq	-0.002*** (0.0003)	-0.001*** (0.0002)	-0.001*** (0.0002)
age	0.0004 (0.003)	0.0002 (0.002)	-0.0004 (0.002)
marriedsingle	-0.022 (0.018)	-0.027** (0.014)	-0.017 (0.013)
Constant	0.534*** (0.146)	0.708*** (0.116)	0.608*** (0.114)
Observations	1,669	2,229	2,229
R ²	0.208	0.147	0.264
Adjusted R ²	0.206	0.145	0.262
Residual Std. Error	0.344 (df = 1662)	0.308 (df = 2222)	0.303 (df = 2222)
F Statistic	72.917*** (df = 6; 1662)	63.941*** (df = 6; 2222)	132.791*** (df = 6; 2222)

Note:

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