PS7

Audrey Hopewell

March 10, 2020

1 Question 6

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,686	1.622	0.388	0.005	1.358	1.936	2.261
hgc	2,244	13.099	2.521	0.000	12.000	15.000	18.000
tenure	2,231	5.978	5.510	0.000	1.583	9.333	25.917
age	2,246	39.153	3.060	34	36	42	46

Almost 25% of log wage observations are missing. They are probably missing not at random, because I would guess that women with lower wages (as related to fewer years of school or not graduating from college) would be less likely to report wage.

2 Question 7

Table 2: Results

	Dependent variable:		
	logwage		
	(1)	(2)	(3)
hgc	0.062***	0.049***	0.062***
	(0.005)	(0.004)	(0.004)
collegenot college grad	0.146***	0.160***	0.146***
	(0.035)	(0.026)	(0.025)
tenure	0.023***	0.015***	0.023***
	(0.002)	(0.001)	(0.001)
age	−0.001	−0.001	−0.001
	(0.003)	(0.002)	(0.002)
marriedsingle	−0.024	−0.029**	−0.024*
	(0.018)	(0.014)	(0.013)
Constant	0.639***	0.833***	0.639***
	(0.146)	(0.115)	(0.111)
Observations	1,669	2,229	2,229
R^2	0.195	0.132	0.268
Adjusted R^2	0.192	0.130	0.266
Residual Std. Error	0.346 (df = 1663)	0.311 (df = 2223)	0.300 (df = 2223)
F Statistic	80.508*** (df = 5; 1663)	67.496*** (df = 5; 2223)	162.884*** (df = 5; 2223)
Note:			*p<0.1; **p<0.05; ***p<0.01

3 Question 8

For my project, I'm using a Kaggle dataset that uses data from IMDb. It includes information on movie title, genre, production company/country, spoken languages, runtime, budget, revenue, IMBDb rating and number of ratings, and cast/crew. I'm thinking about doing some sentiment analysis of movie titles by genre, and perhaps looking at relationships between genre/budget/ratings/creative team for a number of specific actors (along the lines of what FiveThirtyEight did for Nicolas Cage movies here).