

Problem Set 7

Mason Ross Hayes

March 2020

1 Wages

Table 1: Using the stargazer package on a tibble results in null output

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
-----------	---	------	----------	-----	----------	----------	-----

Table 2: stargazer must receive a standard data frame as input and not a tibble

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

6. The variable logwage is missing for 25% of the observations. It is most likely MNAR; those with a very low or very high wage may not wish to report their incomes; or, some other factor could explain why people wish not to report their incomes, such as age for example.

2 Imputations

Table 3: Comparing the three regression models

	<i>Dependent variable:</i>		
	Complete	logwage Mean Imputed	LM Imputed
	(1)	(2)	(3)
hgc	0.062*** (0.005)	0.050*** (0.004)	0.048*** (0.004)
collegenot college grad	0.145*** (0.034)	0.168*** (0.026)	0.174*** (0.026)
tenure	0.050*** (0.005)	0.038*** (0.004)	0.037*** (0.004)
tenure_sq	-0.002*** (0.0003)	-0.001*** (0.0002)	-0.001*** (0.0002)
age	0.0004 (0.003)	0.0002 (0.002)	0.0003 (0.002)
marriedsingle	-0.022 (0.018)	-0.027** (0.014)	-0.029** (0.014)
Constant	0.534*** (0.146)	0.708*** (0.116)	0.720*** (0.117)
Observations	1,669	2,229	2,229
R ²	0.208	0.147	0.131
Adjusted R ²	0.206	0.145	0.129
Residual Std. Error	0.344 (df = 1662)	0.308 (df = 2222)	0.312 (df = 2222)
F Statistic	72.917*** (df = 6; 1662)	63.973*** (df = 6; 2222)	55.966*** (df = 6; 2222)

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

If the true value of $\hat{\beta}_1$ is 0.093, then the mean imputed and lm imputed models are both bad models to use. This is probably because *logwage* is not actually missing at random. Both the mean and lm imputations result in a $\hat{\beta}_1$ that decreases from simply using the complete data.

3 Project

I have made very little progress on the project since I have not received the data from scite yet. I may have to change my project to examine a topic for which there are already good datasets to use. I hope this isn't the case – it depends when I can get the data from scite.