Problem Set 7

Collin DeVore

March 2023

Question 6

logwage is missing 560 observations out of 2,229 observations. In percentage terms, this is 25%. Given that these are the missing values that are left after omitting the missing values for hgc and tenure, it seems likely that these values are missing completely at random. For missing at random, we would need to be able to explain some of the missing values using other variables, which is clearly not the case here. For missing not at random, we would need to be able to have an explanation for why some of the variables are missing, which does not seem to be the case. The NA's appear to be both random and independent of other missing values from other variables, so this is likely missing completely at random.

Question 7

For my values, mean imputation appears to be the most inaccurate predictor. All of the other values are extremely similar. This makes sense, given that one of these methods used predictive values based off of another. One issue I am seeing, however, is that most of these values are more than 0.02 away from the true value. Given that 25% of the values in the dataset are now estimated in different ways, however, it makes sense that these estimates would not be entirely accurate. For the predicted values method, the coefficient estimate is 0.062, which is the same value as for listwise deletion. Multiple implementation is close to this value at 0.061. These values would suggest to me that the true coefficient would be closer to 0.06, so it is odd that these two estimates seem so much lower.

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
logwage	670	25	1.6	0.4	0.0	1.7	2.3
hgc	16	0	13.1	2.5	0.0	12.0	18.0
tenure	259	0	6.0	5.5	0.0	3.8	25.9
age	13	0	39.2	3.1	34.0	39.0	46.0

	Listwise Deletion	Mean Imputation	Predicted Values	Multiple Implementation
(Intercept)	0.639	0.835	0.639	0.659
	(0.146)	(0.115)	(0.111)	(0.136)
hgc	0.062	0.049	0.062	0.061
	(0.005)	(0.004)	(0.004)	(0.005)
collegenot college grad	0.146	0.160	0.146	0.135
	(0.035)	(0.026)	(0.025)	(0.029)
tenure	0.023	0.015	0.023	0.023
	(0.002)	(0.001)	(0.001)	(0.002)
age	−0.001	−0.001	−0.001	−0.001
	(0.003)	(0.002)	(0.002)	(0.003)
marriedsingle	−0.024	−0.029	−0.024	−0.020
	(0.018)	(0.014)	(0.013)	(0.018)
Num.Obs.	1668	2220	2220	2220
Num.Imp.				5
R2	0.195	0.132	0.267	0.219
R2 Adj.	0.192	0.130	0.266	0.217
AIC	1206.3	1132.6	966.2	
BIC	1244.3	1172.5	1006.2	
Log.Lik.	−596.161	−559.289	−476.123	
F	80.411	67.422	161.526	
RMSE	0.35	0.31	0.30	

Question 8

I have spoken to Dr. Kim, Dr. Kosmopoulou, and Yue about my project. The data is gathered and I will be meeting with Yue to get it tomorrow. Basically, we will run regressions to estimate the closure rate of small businesses during the pandemic. Yue and I will be working on this at the same time, and we will be combining anything interesting that we find. We have county level vaccination data, which may mean that our basic panel regression model may adapt into a different-in-difference style of regression. Either way, I should be able to turn in a project that has a ton of my own work.