PS7

christopher.p.danko-1

March 2020

1 Problem Set 6

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,671	1.625	0.385	0.005	1.364	1.936	2.261
hgc	2,229	13.101	2.524	0.000	12.000	15.000	18.000
tenure	2,231	5.978	5.510	0.000	1.583	9.333	25.917
age	2,231	39.152	3.061	34	36	42	46

About 1/3 of observations for logwage are missing.
The value of logwage is likely missing at random, though it could be missing not at random if people are embarrassed about giving their wage when it's lower

2 Problem 7

The β_1 for the first regression is pretty similar to the multiple imputation coefficient, and they are both the closest coefficients to the true value without missing data. Both mean imputation and predictive mean imputation are pretty far from the expected value. The estimates for single imputation and multiple imputation are very similar, and have very similar standard errors. (Table 2)

3 Problem 8

I have a dataset, and I've started to work on the code for a model. The model I'm using is a collection of college basketball matchups, with season ratings for both teams and the score difference. I would like to do something complicated like a neural net, but I may settle on something simpler.

Table 2: Regression Results

	Dependent variable:			
	Log Wage			
	(1)	(2)	(3)	(4)
HGC	0.062^{***}	0.049^{***}	0.057^{***}	0.061^{***}
	(0.005)	(0.004)	(0.005)	(0.005)
College	0.144^{***}	0.168^{***}	0.130^{***}	0.158^{***}
	(0.034)	(0.026)	(0.029)	(0.028)
Tenure	0.050^{***}	0.038^{***}	0.055^{***}	0.046^{***}
	(0.005)	(0.004)	(0.004)	(0.004)
Tenuresq	-0.002^{***}	-0.001^{***}	-0.002^{***}	-0.001^{***}
	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Age	0.001	0.0003	−0.00000	−0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Married	−0.024	-0.028^{**}	−0.024	−0.012
	(0.018)	(0.014)	(0.015)	(0.015)
Constant	0.531^{***}	0.707^{***}	0.623^{***}	0.578^{***}
	(0.146)	(0.116)	(0.130)	(0.126)
Observations	1,660	2,219	2,231	2,231
R^2	0.210	0.147	0.213	0.240
Adjusted R^2	0.207	0.145	0.210	0.238
Residual Std. Error	0.343 (df = 1653)	0.308 (df = 2212)	0.347 (df = 2224)	0.335 (df = 2224)
F Statistic	73.222^{***} (df = 6; 1653)	63.648^{***} (df = 6; 2212)	100.068^{***} (df = 6; 2224)	116.769^{***} (df = 6; 2224)
Note:				*p<0.1; **p<0.05; ***p<0.01