PS7Buchman

blakebuchman03

March 2021

1 Questions

6. At what rate are log wages missing? Do you think the logwage variable is is most likely to be MCAR, MAR, or MNAR?
Logwage was missing in 25 percent of the data after we took out missing variables for hgc and tenure. I believe the missing data points are MAR. The observations that are missing become a lot less common as hgc gets lower. For example, logwage is there at a much higher frequency at people with hgc of less than 10 then it is above 15.

7. Beta 1 is listed as hgc in the formula, the coefficient for hgc is .062, .050, .062 and .062 so it was very consistent except for Model 2. It was not very accurate and I am not sure how to fix that, but it was consistently off from the .093 that it was supposed to be. The last two are consistent with each other.

8. I have figured out what I'm going to do on my project and have a plan on how to accomplish it. I am using college basketball data off of Kaggle that shows me various team statistics from this season for every team. I am going to build a model that tries to predict the outcome of college basketball NCAA tournament games using statistics from that teams season. First I have to merge multiple data sets and figure out how to get everything lined up so I can do the model.

	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
logwage2	670	0	1.6	0.3	0.0	1.6	2.3
logwage3	1221	0	1.7	0.3	0.0	1.7	2.3

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.534	0.708	0.534	0.534
	(0.146)	(0.116)	(0.112)	(0.146)
hgc	0.062	0.050	0.062	0.062
	(0.005)	(0.004)	(0.004)	(0.005)
as.factor(college)not college grad	0.145	0.168	0.145	0.145
	(0.034)	(0.026)	(0.025)	(0.034)
poly(tenure, 2, raw = T)1	0.050	0.038	0.050	0.050
	(0.005)	(0.004)	(0.004)	(0.005)
poly(tenure, 2, raw = T)2	-0.002	-0.001	-0.002	-0.002
	(0.000)	(0.000)	(0.000)	(0.000)
age	0.000	0.000	0.000	0.000
	(0.003)	(0.002)	(0.002)	(0.003)
as.factor(married)single	-0.022	-0.027	-0.022	-0.022
	(0.018)	(0.014)	(0.013)	(0.018)
Num.Obs.	1669	2229	2229	1669
Num.Imp.				5
R2	0.208	0.147	0.277	0.208
R2 Adj.	0.206	0.145	0.275	0.206
AIC	1179.9	1091.2	925.5	
BIC	1223.2	1136.8	971.1	
Log.Lik.	-581.936	-537.580	-454.737	
F	72.917	63.973	141.686	