

Problem Set 6

Leah Pomerantz

March 29, 2022

1 Problem 6

The table produced by R is shown below. logwages are missing at a rate of 25%. With such a high percentage missing, I would guess that it's MNAR.

2 Problem 7

The table produced by R with all the regression results is shown below. The $\hat{\beta}_1$ estimates vary. from 0.050 up to 0.062. What is consistent about all the values is that they have hugely underestimated the value of $\hat{\beta}_1$. However, out of all the estimations, the listwise deletion and predicted values got the closest. This suggests that the data may be MCAR or MAR. It's a reminder that different methods work better/worse for different data sets, depending on the reason why the data are missing. The last two values for $\hat{\beta}_1$ are close together, suggesting that the methods are both working okay (at least working better than mean imputation), but the predicted values method still does a slightly better job at the prediction.

3 Problem 8

For my project, I plan on using data from my MA thesis. Right now, I'm thinking about looking at how having a parent's education level affects a child's education level. For example, I would expect a high school graduate mother to have a negative affect on probability of going to college, and I would expect the difference between some college and college graduate to be smaller. This prediction comes from my own experience working in admissions and recruitment. There's a steep knowledge gap in how to get a child into college, from things as simple as taking the SATs and PSATs, to taking APs, Honors classes, etc. I think having some college means that the parent will know more about how to get those things accomplished, even though they don't have a degree. That said, I only have data I can match for mothers, not fathers, so that would be a huge flaw with the model.

I think I would use an ordered logistic regression for this. I have an ordered categorical variable as my dependent variable. I've never done this type of analysis, though, so I would need to read up on it some more.

	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	Mice
(Intercept)	0.534 (0.146)	0.708 (0.116)	0.534 (0.112)	0.638 (0.155)
hgc	0.062 (0.005)	0.050 (0.004)	0.062 (0.004)	0.058 (0.005)
collegenot college grad	0.145 (0.034)	0.169 (0.026)	0.145 (0.025)	0.109 (0.029)
tenure	0.050 (0.005)	0.038 (0.004)	0.050 (0.004)	0.043 (0.005)
I(tenure^2)	-0.002 (0.000)	-0.001 (0.000)	-0.002 (0.000)	-0.001 (0.000)
age	0.000 (0.003)	0.000 (0.002)	0.000 (0.002)	0.000 (0.003)
marriedsingle	-0.022 (0.018)	-0.027 (0.014)	-0.022 (0.013)	-0.020 (0.015)
Num.Obs.	1669	2229	2229	
R2	0.208	0.146	0.277	
R2 Adj.	0.206	0.144	0.275	
AIC	1179.9	1093.8	925.5	
BIC	1223.2	1139.5	971.1	
Log.Lik.	-581.936	-538.912	-454.737	
F	72.917	63.461	141.686	
RMSE	0.34	0.31	0.30	