PS7

Kyle Davis

March 26, 2024

1

The rate at which logwages is missing is 0.69 percent and is MCAR.

2

The models generated the same result on every variable which I imagine isn't supposed to happen but AI said it was just because I was using the same variables so I am unsure. They all equaled 0.062 for the coefficient. The pattern is they all are the same which seems off as I mentioned. The selection of an imputation method depends on the nature of missing data, data complexity, and the desired balance between efficiency and robustness. The last two estimates were identical to the former as well as with one another.

3

My project is finished as far as the data. I am using statistics from 2021-2023 to create a prediction of 2024 fantasy score using variables that are more consistent across someones career such as hard hit rate, contact rate, and home run to fly ball ratio. I am doing it over three years to create adjustments for each player based on a moving average and resulting in a fantasy point per game number that I used to draft my team this year. I utilized a lasso regression since there is some intercorrelation between the independent variables.

	Complete Cases	Mean Imputed	Predicted	Multiple Imputed
(Intercept)	0.534	0.534	0.534	0.534
	(0.146)	(0.146)	(0.146)	(0.146)
hgc	0.062	0.062	0.062	0.062
	(0.005)	(0.005)	(0.005)	(0.005)
collegenot college grad	0.145	0.145	0.145	0.145
	(0.034)	(0.034)	(0.034)	(0.034)
tenure	0.050	0.050	0.050	0.050
	(0.005)	(0.005)	(0.005)	(0.005)
tenure2	−0.002	−0.002	−0.002	−0.002
	(0.000)	(0.000)	(0.000)	(0.000)
age	0.000	0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
marriedsingle	−0.022	−0.022	−0.022	−0.022
	(0.018)	(0.018)	(0.018)	(0.018)
Num.Obs.	1669	1669	1669	1669
Num.Imp.				5
R2	0.208	0.208	0.208	0.208
R2 Adj.	0.206	0.206	0.206	0.206
AIC	1179.9	1179.9	1179.9	
BIC	1223.2	1223.2	1223.2	
Log.Lik.	−581.936	−581.936	−581.936	
F	72.917	72.917	72.917	
RMSE	0.34	0.34	0.34	

Figure 1: Table.