

# PS7 Dawkins

Julie Dawkins

March 2024

## 1 Problem 6

Here is the table produced by the package modelsummary:

	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

25% of logwage is missing. I believe this is most likely to be Missing Not at Random (MNAR). Wages are sensitive topics and individuals are not likely to randomly not report them. Although we can hypothesize their values based on other variables (age, tenure, education, etc.) we cannot fully know these values based on these variables.

## 2 Problem 7

The following is the result of the four types of regressions:

	List-Wise Deletion	Mean Imputation	Predicted Values	Mice
(Intercept)	0.639*** (0.146)	0.833*** (0.115)	0.639*** (0.111)	0.724*** (0.135)
hgc	0.062*** (0.005)	0.049*** (0.004)	0.062*** (0.004)	0.059*** (0.006)
collegenot college grad	0.146*** (0.035)	0.160*** (0.026)	0.146*** (0.025)	0.106** (0.033)
tenure	0.023*** (0.002)	0.015*** (0.001)	0.023*** (0.001)	0.022*** (0.002)
age	-0.001 (0.003)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
marriedsingle	-0.024 (0.018)	-0.029* (0.014)	-0.024+ (0.013)	-0.019 (0.018)
Num.Obs.	1669	2229	2229	
R2	0.195	0.132	0.268	
R2 Adj.	0.192	0.130	0.266	
AIC	1206.1	1129.3	961.2	
BIC	1244.0	1169.3	1001.1	
Log.Lik.	-596.049	-557.651	-473.584	
F	80.508	67.496	162.884	
RMSE	0.35	0.31	0.30	

+  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

None of the  $\hat{\beta}_1$ s come very close to the true value of 0.093. However, List-Wise Deletion and Predicted Values come closer at 0.062, indicating a 6.2% increase in wage for each additional year of schooling. Mean Imputation was the farthest with a  $\hat{\beta}_1$  of 0.049, and the mice multiple imputation method was 0.059.

Based on the true value of  $\hat{\beta}_1$ , I conclude that list-wise deletion and using predicted values were the most accurate for this particular data set. They have the closest  $\hat{\beta}_1$ , and have the same *betas* for every other variable (with differing standard errors).

The  $\hat{\beta}_1$  for the predicted-value method is 0.062, the same as in list-wise deletion; for the mice method, it is 0.059. These are very similar values. The

other *betas* differ more dramatically, with college graduation having a lower impact in the mice method than the others as well as being married (though this variable is not significant in the mice method).

### 3 Problem 8

I have not made it very far on my project. Though I downloaded data for the last two problem sets, I don't think I will be using the same data, as the easiest way to access robust data on the topic of homelessness/housing is to download excel files publicly available by HUD.

I will be looking at homelessness point-in-time counts (conducted by Continuums of Care, regional organizations that monitor homelessness); housing inventory counts; and median rent in a region, all gathered from Excel files from HUD. AmericanCommunities.Org has a map of 171 counties that have major universities; I could compile data from this list to get a more robust data set than the 9 that I used in my last dataset. It will just be a grind to get the relevant info, pulling information about college enrollment,

I don't know if I will find anything significant or even major differences in levels of homelessness between college-town counties and others, but I am interested in exploring the impact of students on homelessness and rents. I could conduct a simple OLS regression as well with either the number of homelessness or rents as my dependent variable.