

PS7

Christin Bivens

March 13th, 2018

- Part 6

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
logwage	1,669	1.625	0.386	0.005	2.261
hgc	2,229	13.101	2.524	0	18
tenure	2,229	5.971	5.507	0.000	25.917
age	2,229	39.152	3.062	34	46

Missing logwage records make up 25.12337 percent of the logwage inputs. The missing records are probably of type MNAR. At first glance, knowing these are women, the logwage missing values might be attributed to those women being housewives. However, the tenure values vary within the missing records. One woman who has a missing record is 37, single, with a tenure of 5.33. It is unlikely that a woman who makes the choice not to work would have a 5 year experience of doing so and be single. The other explanation for this could be that she is unemployed and has been for 5.33 years. Using priors/a Bayesian approach to this problem, it is socially acceptable and normalized for women to be unemployed. The percentage of missing logwage entries is actually smaller than the Federal Reserve's reports on women not participating in the labor force.

- Part 7

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
Using Mean	2,229	1.625	0.334	0.005	2.261
Using LM	2,229	1.635	0.339	0.005	2.261
Using MICE	2,229	1.787	0.474	0.005	3.010

I clearly had no idea what was going on with most of the bullets. I had a fever until this morning and have been trying to figure this out over the

weekend. I should have emailed you sooner. How does doing a simple mean imputation create a regression model from bullet 2? How are bullet 1 and 3 different?

So I made an output table of summary stats for my logwage vectors resulting from mean imputation, using `lm()`, and using MICE. I understand I will not get a high grade on this assignment.

item8

My final project is coming along well. I have updated my data set to have a cleaned Price column. Now I am creating summary tables and basic linear regressions for the variables of interest, as well as simple correlation tests. I will most likely be doing a multivariate regression. The independent variables and coefficients of interest will be decided by using insights from the work I am doing now.