

# PS7 Stinnett

sloan.c.stinnett-1

March 2019

Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,669	1.625	0.386	0.005	1.362	1.936	2.261
hgc	2,229	13.101	2.524	0	12	15	18
tenure	2,229	5.971	5.507	0.000	1.583	9.333	25.917
age	2,229	39.152	3.062	34	36	42	46

## 1 rate of missing values in logwages

out of 2229 observations 560 are missing values for log wages. this means that approximately 25 percent of all observations are missing logwage data. It is most likely that the values are missing in a non-random manner.

## 2 differences in the effect of hgc across the four methods

methods 1 and 4 gave similar predictions both of which where better predictions of the true value than those of method 2 and 3. two of the three imputation methods gave poorer estimates than the deletion method. It would seem that the mean imputation method and the regression imputation method both give poor results and would fall outside of best practice. the method used by MICE and the deletion method both seem to be reliable. The last  $\beta_1$ s where 0.048329 for method 3 and 0.06169 for method 4.

## 3 Progress on project

I have found some interesting tools that will allow me to gather a comprehensive data set. The fist of these is PACER a comprehensive library of district and

appellate court documents. The second is a tool for python called juriscraper which is a tool to scrape files from PACER into python. I am still figuring out the scope I want the project to have I will then begin gathering the data set.