

PS7 - ECON 5253

Hannah Bermudez

February 2025

1 Summary using Modelsummary Package

Variable	Unique	Missing Pct.	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
N		%					
College	College grad	530	23.8				
	Not college grad	1699	76.2				
Married	Married	1431	64.2				
	Single	798	35.8				

Table 1: Summary statistics of variables

1.1 logwage: MCAR, MNAR, MAR?

The logwage variable is missing 25.12 percent of the observations.

This suggests that the missing data for ‘logwage’ is likely Missing Not at Random (MNAR) or Missing at Random (MAR), given that 25 percent of the observations are missing. The nonresponse may be influenced by factors related to the data itself, such as wage level, where individuals with particularly high or low wages might choose not to report their earnings due to concerns about judgment. Additionally, missingness may be associated with education or other systemic factors contributing to nonresponse. As a result, the missing data cannot be classified as Missing Completely at Random (MCAR).

2 Regressions

	Listwise Deletion	Mean Imputation	Regression Imputation	Multiple Imputation
(Intercept)	0.534*** (0.146)	0.708*** (0.116)	0.534*** (0.112)	0.630*** (0.162)
hgc	0.062*** (0.005)	0.050*** (0.004)	0.062*** (0.004)	0.059*** (0.006)
collegenot college grad	0.145*** (0.034)	0.168*** (0.026)	0.145*** (0.025)	0.115** (0.033)
tenure	0.050*** (0.005)	0.038*** (0.004)	0.050*** (0.004)	0.044*** (0.004)
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)	-0.017 (0.020)
Num.Obs.	1669	2229	2229	
R2	0.208	0.147	0.277	
R2 Adj.	0.206	0.145	0.275	
AIC	1179.9	1091.2	925.5	
BIC	1223.2	1136.8	971.1	
Log.Lik.	-581.936	-537.580	-454.737	
F	72.917	63.973	141.686	
RMSE	0.34	0.31	0.30	

+ p > 0.1, * p > 0.05, ** p > 0.01, *** p > 0.001

2.1 Patterns/Differences of hgc Beta Across Models

None of the betas for hgc are close to the true beta of 0.093, which suggests that all models exhibit some degree of bias due to the missing data mechanism or the imputation methods used. This could be due to the variation in size between the sample and the population, attributing to sampling error. In addition, the listwise deletion and regression imputation share the same coefficient and similar standard deviations. Considering that the number of observations drops from 2229 to 1669 for the listwise deletion model, using the regression imputation model could be preferable, as it retains more observations while producing an estimate identical to that of listwise deletion. The R-squared for the regression imputation model is also bigger compared to the listwise deletion, proving that the model fit is better and explains a greater proportion of the variability in logwage.

3 Final Project Progress

I am planning to run an experiment with undergraduate and graduate accounting students to look at decision-making in an audit context. Currently, I have sent my instrument to the IRB for approval and I expect to administer the survey sometime near the end of March or beginning of April. I plan to perform ANOVA and regression analysis to examine the effects of certain factors on the revision of

sequential beliefs. Due to my data not being currently available, I have since begun working on the report.