# Problem Set 7

Lilly Kirby-Rivera

March 26, 2024

## Q6.

According to the data summary, *logwage* is missing at a 25% rate. It seems likely that this data is MNAR since the unobserved data is likely to be driven by an unobserved variable. Looking at a correlation table, listed below, we should rule out the possibility of MCAR data since there is a correlation with *hgc* and *tenure*. Looking at the raw data, it seems that those with many years of tenure have higher rates of missing data, so perhaps those with extreme values for income are reluctant to report it.

1. Summary table

| | 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 |

2. Correlation table

| | logwage | hgc | tenure | age | missing |
|---|---|---|---|---|---|
| logwage | 1 | . | . | . | . |
| hgc | 0.30 | 1 | . | . | . |
| tenure | 0.33 | 0.12 | 1 | . | . |
| age | −0.01 | −0.03 | 0.07 | 1 | . |
| missing | | 0.37 | 0.23 | −0.01 | 1 |

## Q7.

Listwise deletion and the predicted values models are the closest to the true parameter of 0.093. Mean imputation yields the poorest result by estimating it to be 0.05 while multiple imputation is quite close to the other two models. Since the model for predicted values is a much worse estimate compared to the true value, the data is unlikely to be MAR. This demonstrates that missing observations can have an extreme impact on the explanatory power of your model, even when correcting for missing data with various methods. For this data, it is likely that the true parameter values are driven by the variation of outliers or more extreme values of *logwage* that are under-reported.

3. Regression table

| | Listwise deletion | Mean imputation | Predicted values | Multiple imputation |
|---|---|---|---|---|
| (Intercept) | 0.534*** | 0.708*** | 0.534*** | 0.670*** |
| | (0.146) | (0.116) | (0.112) | (0.142) |
| hgc | 0.062*** | 0.050*** | 0.062*** | 0.059*** |
| | (0.005) | (0.004) | (0.004) | (0.006) |
| collegenot college grad | 0.145*** | 0.168*** | 0.145*** | 0.111** |
| | (0.034) | (0.026) | (0.025) | (0.036) |
| tenure | 0.050*** | 0.038*** | 0.050*** | 0.041*** |
| | (0.005) | (0.004) | (0.004) | (0.005) |
| tenure2 | −0.002*** | −0.001*** | −0.002*** | |
| | (0.000) | (0.000) | (0.000) | |
| age | 0.000 | 0.000 | 0.000 | −0.001 |
| | (0.003) | (0.002) | (0.002) | (0.002) |
| marriedsingle | −0.022 | −0.027* | −0.022+ | −0.013 |
| | (0.018) | (0.014) | (0.013) | (0.016) |
| I(tenure^2) | | | | −0.001*** |
| | | | | (0.000) |
| Num.Obs. | 1669 | 2229 | 2229 | 2229 |
| Num.Imp. | | | | 5 |
| R2 | 0.208 | 0.147 | 0.277 | 0.215 |
| R2 Adj. | 0.206 | 0.145 | 0.275 | 0.213 |
| AIC | 1179.9 | 1091.2 | 925.5 | |
| BIC | 1223.2 | 1136.8 | 971.1 | |
| Log.Lik. | −581.936 | −537.580 | −454.737 | |
| RMSE | 0.34 | 0.31 | 0.30 | |

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

# Q8.

For my project, I am planning to estimate the entry deterring power of fraudulently listed patents in the FDA Orange Book. This research question is of recent interest to the FTC, who demanded in December 2023 that 10 manufacturers remove what they suspect to be improperly listed patents from the Orange book. While they hypothesize that improper patent listings may dissuade generic manufacturers from entering the market by imposing greater legal costs, there are no formal studies that provide estimates on their impact. I have already done a literature review and written up the background section of my paper. I am planning to use the Medicare Part D dataset for this paper, but I applied recently for the DFCAS dissertation research fellowship so I may purchase higher quality proprietary data. For this project, it will be essential to have accurate and rich data on price and quantity for the treatment and control groups of pharmaceuticals. I plan to use the probit model to estimate the likelihood of generic entrance for some number of periods with and without improper patent listing (by product).