To: Professor Yuce

Applied Mathematics

New York City College of Technology

From: Brian Holliday

Applied Mathematics

New York City College of Technology

Subject: Lab 6: MACROS and FILE I/O

Date: 2/13/20

Encl:

Ref: Figure 1: hs\_wfdata

Summary:

In this **hs\_wfdata**, we have some missing data hindering our analysis. If missing data is not taken care of it propagates throughout the analysis, making our new datasets have less real values. We could see this in the **wife age at marriage** and **year marriage** columns. To fill missing values, if there is a small standard deviation, we could put the mean. If there is a large standard deviation, put the median. If we have a large amount of repeating numbers, then we could use the mode. We could also use a regression model to fill in for missing data, coming up with a curve that fits our data in order to predict missing data values and future values. In this case to fill the **wife age** column, I would take the median as the data seems to have a significant variance in that column. With the **husband age** at marriage, I would probably just insert the mean as all of the data is in the 20’s. Sometimes a combination of all the methods is best.

