To: Professor Yuce

Applied Mathematics

New York City College of Technology

From: Brian Holliday

Applied Mathematics

New York City College of Technology

Subject: Lab 14: Modeling, missing data values and data imputation

Figure:

Figure 1: ha wa MEANS Procedure

Figure 2: ha wa correlation

Figure 3: hh wh correlation

Figure 4: Regression hh wh

Figure 5: hh wh scatter

Figure 6: Cubic regression

Date: 3/12/20

Summary:

In this lab we are trying to do some basic statistics for our dataset for our husband and wife dataset. We will look at mean and standard deviation of some of our variable, along with trying to find the correlation between variables.

3.

![A screenshot of a cell phone

Description automatically generated]()

We see that husband age has 199 values, zero missing values, a mean age of 42.62, and standard deviation of 11.64. We see that wife age has 170 values with 29 missing values, a mean of 40.68 anstandard deviation of 11.41

4.

![A screenshot of a cell phone

Description automatically generated]()

There is a strong correlation between husband age and wife age at 0.9385.

![A screenshot of a cell phone

Description automatically generated]()

The is a low correlation between husband height and wife height at 0.3644.

5.

![A screenshot of a cell phone

Description automatically generated]()

We see that the R^2 score for hh and wh is 0.1328, meaning that hh is not a great indicator for wh. I feel that this is a consistent with the correlation score of 0.3644. This means that where is not a strong relationship between these variables.

![A screenshot of a cell phone

Description automatically generated]()

From the graph, we could see there is a slight positive correlation between husband height and wife height. Based on this graph I do not think that a quadratic model would be better, there are a lot of outliers, but let’s try it.

![A screenshot of a cell phone

Description automatically generated]()

We see that for husband age and wife age, the p value tells us that the coefficients for these variables are too close to zero for them to be significant. The linear model is better.