Homework 5 – Part 1

1. Using Davis's data on measured and reported height and weight do the following[1]:

   - Construct a dummy regressor for sex, and regress reported weight on measured weight and sex. Form the regression equations for the two groups and interpret each of the regression coefficients in the two equations.
   - Note that observation 12's measured weight and height were flipped when the data was entered. Fix this error and reinterpret the model.

Table 1: Linear Regression of Reported Weight

	Estimate	Std. Error
Gender (Male=1)	9.751	1.177
Weight (in kg.)	0.570	0.038
Constant	23.520	2.328

$$\widehat{ReportedWeight} = 23.520 + 9.751(Gender) + 0.570(Weight)$$

Males (Gender=1)

$$\widehat{ReportedWeight} = 23.520 + 9.751(1) + 0.570(Weight)$$
$$\widehat{ReportedWeight} = 33.271 + 0.570(Weight)$$

Females (Gender=0)

$$\widehat{ReportedWeight} = 23.520 + 9.751(0) + 0.570(Weight)$$
$$\widehat{ReportedWeight} = 23.520 + 0.570(Weight)$$

Males' reported weights are on average 9.751 kilograms higher than females, ceteris paribus. Holding gender constant, a kilogram increase in measured weight averages a 0.570 kilogram increase in reported weight.

[1] This data is available in R, via the car library.

Table 2: Linear Regression of Reported Weight, Error Fixed

	Estimate	Std. Error
Gender (Male=1)	1.563	0.467
Weight (in kg.)	0.974	0.017
(Intercept)	1.019	1.021

$$\widehat{ReportedWeight} = 1.019 + 1.563(Gender) + 0.974(Weight)$$

Males (Gender=1)

$$\widehat{ReportedWeight} = [1.019 + 1.563(1)] + 0.974(Weight)$$
$$\widehat{ReportedWeight} = 2.582 + 0.974(Weight)$$

Females (Gender=0)

$$\widehat{ReportedWeight} = [1.019 + 1.563(0)] + 0.974(Weight)$$
$$\widehat{ReportedWeight} = 1.019 + 0.974(Weight)$$

After fixing the data entry error, this model seems to make a little more sense. Now, holding gender constant, a kilogram increase in measured weight averages a 0.974 kilogram increase in reported weight. This is much closer to the one-to-one tracking ($B_2 = 1$) that you would expect if people were accurately reporting their weight. The gender difference is more muted now. Females, on average, have reported weights that are 1.563 kilograms lower than males of the same weight.

2. The following analysis was estimated using data on the amount of exercise completed (in hours per week) by girls in two groups. One group of girls have been hospitalized for eating-disorders (Patient = 2). The other is a control group, none of which have been hospitalized for eating disorders (Patient = 1). The age of each subject is also recorded. An expert in eating-disorders has estimated the following equation, but needs your help interpreting the results.

$$EXE\hat{R}CISE = -2.72 + 1.31(PATIENT) + 0.68(AGE)$$

Substantively interpret this model. What is the difference between girls with and without eating disorders? Please note that the expert used 1 and 2 as the values for the dummy regressor, rather than the standard 0 and 1. Interpret the results accordingly. Drawing a graph may be helpful.

Hospitalized Patients (PATIENT=2)

$$EXE\hat{R}CISE = [-2.72 + 1.31(2)] + 0.68(AGE)$$
$$EXE\hat{R}CISE = -0.10 + 0.68(AGE)$$

Non-hospitalized Patients (PATIENT=1)

$$EXE\hat{R}CISE = [-2.72 + 1.31(1)] + 0.68(AGE)$$
$$EXE\hat{R}CISE = -1.41 + 0.68(AGE)$$

Just like a model that had been code with (0,1), the difference between the groups is represented by $B_1 = 1.31$. This is verified by plugging 2 and 1 into predicted exercise equations and simplifying. The difference in the intercept for group 1 and group 2 ($-0.10 - -1.41 = 1.31$) shows the difference between group 1 and 2. Girls that have been hospitalized for eating disorders exercise an average of 1.31 hours more per week than girls that have not been treated for eating disorders, ceteris paribus.

If the difference between group 1 (Patient=2) and group 2 (Patient=1) had been greater or less than 1, B_1 alone would not represent the group difference. Only by plugging the specific values used to signify groups into the equations and simplifying, as we did above, can the differences between groups be positively determined.