1. (2 points) What happens to our OLS inference if we *multiply* some value by X, where X is our explanatory variable of interest?

2. This question relies on the following file: anorexia.csv. This data examines the effectiveness of different treatments for anorexia.
   (a) (3 points) Treat the data as paired data (as it is). I'm interested in whether treatment "f" is effective in adding weight to the patients. Derive a test statistic to determine a difference. What is your decision with the null based on the test?
   (b) (2 points) Why is it particularly difficult to infer a causal effect using a "matched pair" test?
   (c) (2 points) Repeat the same test procedure for treatment "c", Derive a test statistic to determine a difference. What is your decision with the null based on the test?
   (d) (2 points) Repeat the same procedure for treatment "b", Derive a test statistic to determine a difference. What is your decision with the null based on the test?
   (e) (2 points) We just engaged in some "multiple testing" behavior. What is the potential problem of multiple testing? Propose a strategy to address this "multiple testing" problem (you don't have to actually carry this out - just discuss it). Is there a tradeoff with your strategy?
   (f) (2 points) Now suppose we failed to recognize that the data were paired. Derive a difference in means test statistic for the "b" groups (assume unequal variance).
   (g) (2 points) Was unequal variance a justifable assumption for the previous test. Use an F-test to support your case.

3. This next section will require you to use the following dataset: "gerber_green_larimer.Rdata" (this is the social pressure/civic duty data from the slides, weeks 6 and 7).
   (a) (2 points) Present a descriptive graph that compares voting turnout for each treatment group.
   (b) (1 point) For now, assume that variable "voted" is normally distributed. Run a one-way ANOVA with "treatment" as the main explanatory variable. What is the null hypothesis of this test?
   (c) (1 point) What is the sampling distribution of this test? Please identify the degrees of freedom and where the sampling distribution is centered.
   (d) (1 point) What is the explained and unexplained variance of your ANOVA test?
   (e) (1 point) What's the F-statistic derived from the ANOVA test you ran? What does it tell us from a statistical significance standpoint?
   (f) (2 points) Reconsider the outcome variable (voted). Is an ANOVA appropriate with these data?

4. This next section will require you to use the following dataset: midtermvoteloss.csv.

(a) (2 points) Present a scatter plot of Midterm Vote Loss (the dependent variable) and Change in Income (the independent variable). Show this scatter plot with the best fitted line.

(b) (2 points) Estimate the linear model associated with this scatter plot. Present the results and interpret all the important information.

(c) (2 points) Provide two point predictions with the linear model (i.e. with minimum value of Change in Income and maximum value of Change in Income, etc.).

5. This next section will require you to use the following dataset: hw7.

(a) (2 points) Ignoring what the variables are for a moment, estimate the following linear model using OLS and report the results:

$$y1 = \alpha + \beta x1 + \epsilon$$

(b) (2 points) Estimate the following linear model using OLS and report the results:

$$y2 = \alpha + \beta x2 + \epsilon$$

(c) (2 points) Estimate the following linear model using OLS and report the results:

$$y3 = \alpha + \beta x3 + \epsilon$$

(d) (2 points) Estimate the following linear model using OLS and report the results:

$$y4 = \alpha + \beta x4 + \epsilon$$

(e) (3 points) Given what you have learned in class, how would you compare the linear models you just estimated?