# Problem Set 2

Statistics and Econometrics

*Due: 11am, 11 November 2020*

## General Guideline

What we are looking for in the assignments is a demonstration that you can understand the econometrics and statistics questions and can solve them with R or conceptually. That means effective programming to get correct results is needed, but at the same time, clear explanations of economics/business concepts in well presented reports are equally important when assessing your work. In particular, you will be marked for successful (correct) programming (not the style of coding), good understanding of related concepts, and clear interpretations and explanations of results.

Please submit a pdf or html file converted from R markdown/notebook after you program in R.

## Question 1

Use bwght.RData for this exercise. The data set contains data on births to women in the United States. A problem of interest to health officials (and others) is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight (variable *bwght*); a birth weight that is too low can put an infant at risk for contracting various illnesses. Since factors other than cigarette smoking (variable *cigs*) that affect birth weight are likely to be correlated with smoking, we should take those factors into account. For example, higher income (variable *faminc*) generally results in access to better prenatal care, as well as better nutrition for the mother. An equation that recognizes this is:

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 faminc + u.$$

1. What is the most likely sign for $\beta_2$?

2. Do you think *cigs* and *faminc* are likely to be correlated? Explain why the correlation might be positive or negative.

3. Now, estimate the equation with and without *faminc*. Report the results. Discuss your results, focusing on whether adding *faminc* substantially changes the estimated effect of *cigs* on *bwght*.

## Question 2

The following model can be used to study whether campaign expenditures affect election outcomes:

$$voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u,$$

where *voteA* is the percentage of the vote received by Candidate A, *expendA* and *expendB* are campaign expenditures by Candidate A and B, and *prtystrA* is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

1. What is the interpretation of $\beta_1$?

2. Estimate the given model using vote1.RData and report the results. Do A's expenditures affect the outcome? What about B's expenditures? (please show clearly the test statistic and the critical value used in your testing).

3. In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

4. Test the hypothesis in part 3. What is your conclusion?

## Question 3

In class, we discussed the concepts of individual significance (Slide 15) and joint significance (Slide 36). Is it possible that a variable is individually significant but when we test the joint significance of this variable, along with some other variables, we find them jointly insignificant? Explain.