Programming Exercises for Lecture 4

Submission deadline: **June-04-2024**

1. Examine the following statement about a confidence interval (CI) by using a simulation:

   If random samples were drawn again and again, with a 95% (or 99%) CI computed each time, the (unknown) population parameter β_j would lie in these CI's over 95% (or 99%) of the time.

   If the true regression in population is:

$$y = 0.4 - 1.46x_1 + 2.5x_2 + u, \qquad (1)$$

   where $u \sim N(0, 1)$, $x_1 \sim U(0, 1)$ and $x_2 \sim N(0.78, 1)$.

   a. Simulate 300 samples of (x_{1i}, x_{2i}) and u_i and generate y_i with (1). Run a regression of y on (x_1, x_2) with the OLS. Calculate the 95% and 99% confidence intervals for the parameters. Repeat the procedure 50 times by fixing the samples of (x_{1i}, x_{2i}) but re-generating u_i. What are proportions that these confidence intervals cover true values of the parameters in the 50 simulations? Print out your answer.

   b. Repeat the same simulation in a. but the number of iterations now is 200. Print out your answer.

   c. Repeat the same simulation in a. but the number of iterations now is 1,000. Print out your answer.

2. Use the data in `vote1` to estimate the following linear regression:

$$voteA = \beta_0 + \beta_1 lexpendA + \beta_2 lexpendB + \beta_3 prtystrA + u,$$

   where `voteA` is the percentage of the votes received by Candidate A, `lexpendA` and `lexpendB` are log's of campaign expenditures by Candidates 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). This linear regression can be used to study whether campaign expenditures affect election outcomes.

   a. What are descriptive statistics of the variables used here?

   b. What is the interpretation of β_1?

c. 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.

d. Estimate this linear regression using the data in `vote1`. Do A's expenditures affect the outcome? What about B's expenditures? Use method `f_test` to test the hypothesis in c.

e. Estimate a modified model that gives the t statistic for testing the hypothesis in c. Does the square of the t-statistic equal to the F statistic used in d.? What do you conclude?

f. Test a null against an alternative that $\beta_1 + \beta_2 \neq 1.5$ by using the method `f_test`, and also by estimating a modified model that gives the t statistic (manually calculate it or use the method `t_test`). Compare these results, what do you conclude?

3. Use the data in `hprice1` to estimate the following regression:

$$\log(price) = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

where `price` is the house price measured in thousands of dollars, `sqrft` is square footage of the house, and `bdrms` is the number of bed rooms in the house.

a. What are descriptive statistics of the variables used here?

b. Obtain the percentage change in price when a 150-square-foot bedroom is added to a house (call this percentage change θ_1). What is the estimate of this percentage change in terms of the estimated parameters?

c. To obtain standard error of the percentage change in b., you also can estimate a modified regression model. What is it? Estimate the modified model and obtain the standard error of the percentage change in b. and construct the 95% and 99% confidence intervals. Is the estimate of percentage change in b. statistically significant at $\alpha = 5\%$ and 1%?