Problem Set 1

(Submit through the Hub by 12pm October 19th)

1. The data set ceosal2.RData contains information on chief executive officers for U.S. corporations. Two variables of interest are the annual compensation ($salary$) and the prior number of years as company CEO ($ceoten$).

   (a) Find the average salary and the average tenure in the sample.

   (b) How many CEOs are in their first year as CEO (that is, $ceoten = 0$)? What is the longest tenure as a CEO?

   (c) What is the average salary for CEOs with tenure longer than or equal to the average tenure? What is the average salary for CEOs with tenure shorter than the average tenure?

   (d) Create a graph to examine the relationship between $salary$ and $ceoten$ for all CEOs.

   (e) Estimate the simple regression model

   $$\log(salary) = \beta_0 + \beta_1 ceoten + u,$$

   and report your results (including SRF, the sample size, and R-squared). What is the (approximate) predicted percentage increase in salary given one more year as a CEO?

2. The data set bwght.RData contains data on births to women in the United States. Two variables of interest are the infant birth weight in ounces ($bwght$), and the average number of cigarettes the mother smoked per day during pregnancy ($cigs$).

   (a) Estimate the simple regression model

   $$bwght = \beta_0 + \beta_1 cigs + u,$$

   and report your results (including SRF, the sample size, and R-squared).

   (b) What is the predicted birth weight when $cigs = 0$? What about when $cigs = 20$ (one pack per day)? Comment on the difference.

   (c) Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.

   (d) To predict a birth weight of 125 ounces, what would $cigs$ have to be? Comment.

   (e) The proportion of women in the sample who do not smoke while pregnant is about 85%. Does this reconcile your finding from part (d)?

Important: Please also submit the relevant portions of your log file (delete errant commands and output).