Empirical Exercise: Chapter 4

The data file **Earnings_and_Height**, which contains data on earnings, height, and other characteristics of a random sample of U.S. workers. A detailed description is given in **Earnings_and_Height_Description**, also available in the folder. In this exercise, you will investigate the relationship between earnings and height.

a. What is the median value of height in the sample?
b.
  i.   Estimate average earnings for workers whose height is at most 67 inches.
  ii. Estimate average earnings for workers whose height is greater than 67 inches.
  iii. On average, do taller workers earn more than shorter workers? How much more? What is a 95% confidence interval for the difference in average earnings?
c. Construct a scatterplot of annual earnings (*Earnings*) on height (*Height*). Notice that the points on the plot fall along horizontal lines. (There are only 23 distinct values of *Earnings*). Why? (*Hint:* Carefully read the detailed data description.)
d. Run a regression of *Earnings* on *Height*.
  i.   What is the estimated slope?
  ii. Use the estimated regression to predict earnings for a worker who is 67 inches tall, for a worker who is 70 inches tall, and for a worker who is 65 inches tall.
e. Suppose height were measured in centimeters instead of inches. Answer the following questions about the *Earnings* on *Height* (in cm) regression.
  i.   What is the estimated slope of the regression?
  ii. What is the estimated intercept?
  iii. What is the R^2?
  iv. What is the standard error of the regression?
f. Run a regression of *Earnings* on *Height*, using data for female workers only.
  i.   What is the estimated slope?
  ii. A randomly selected woman is 1 inch taller than the average woman in the sample. Would you predict her earnings to be higher or lower than the average earnings for women in the sample? By how much?
g. Repeat (f) for male workers.
h. Do you think that height is uncorrelated with other factors that cause earning? That is, do you think that the regression error term, has a conditional mean of 0 given *Height* (You will investigate this more in the *Earnings* and *Height* exercises in later chapters.)