Assignment 1

Due: Start of class, Monday, October 12th

Instructions:

Group Work: I encourage all of you to work on the problem set in groups, but each member of the group must write up his or her own solutions and perform his or her own analysis. Problem sets that are verbatim copies will receive a zero.

To submit your solutions, include your answers to these questions and your log file (or other supporting documentation) and submit it to the Assignment on Canvas.

Assignment:

This assignment examines the relationship between housing prices, commuting times and other amenities.

The file ps1_data.dta contains information about the housing prices, number of bedrooms, commuting times and high school test scores for zip codes near Boston, MA.

Variable Name	Description
Zcta5	5-digit zip code
City	City name
Bedroom	Number of bedrooms
MedianPrice	Median home price (calculated by Zillow)
DrivingDistance	Distance in miles to downtown Boston
DrivingTime	Length of morning commute to downtown Boston (Driving)
TransitTime	Length of morning commute to downtown Boston (Public Transit)
MathScore	Fraction of 10^{th} grades scoring "advanced" in Math on State exam
EnglishScore	Fraction of 10^{th} grades scoring "advanced" in English on State exam
ScienceScore	Fraction of 10^{th} grades scoring "advanced" in Science on State exam

1. Open a log file to keep a record of your analysis. Click "File → Log → Begin" and give it a name that will be easy to remember.
2. How many observations are in the dataset? Summarize the data and report the average values for the commuting time variables? What is the longest driving commute to Downtown Boston?

   (Hints: Typing "browse" will let you look at the actual data. Typing "count" will count the observations. Typing "summarize DrivingTime TransitTime" will summarize the two variables.)
3. In class, we illustrated a negative relationship between home prices and commute. Draw a scatter plot of the relationship between the two variables and interpret the scatter plot. **(Hint: when Stata draws the scatter plot, you can copy and paste the scatter plot directly into your solutions.)**
4. Now, run the linear regression to match the scatter plot. Describe how the regression relates to the scatter plot and describe how to interpret the sign and significance of the coefficients. **(Hint: to run the regression, type "regress MedianPrice DrivingTime".)**
5. Now, let's add controls to capture the number of bedrooms. In this regression, the Bedroom controls reflect the increase in MedianPrice associated with 2, 3, 4 and 5, bedrooms relative to single bedroom houses.

Interpret the sign, significance and magnitude of the coefficients on the bedroom controls.

In addition, explain how to interpret the R-squared of the regression. How did the R-squared change with the addition of the bedroom controls? Why?

6. Finally, let's think about adding high school test scores as a measure of school quality. First, draw a scatter plot of the relationship between math scores and driving distance and include a line of best fit.

   What does the scatter plot tells us about the relationship between math scores and driving times?

7. Now run the full regression with DrivingTime, MathScore and the number of Bedrooms.

   Is the coefficient on MathScore of the correct sign? Explain how you would interpret a one unit change in MathScore in this regression.

8. Explain why the coefficient on DrivingTime changed the way it did when MathScore was added to the regression. (Hint: Look back at your scatter plot from question 6.)

9. Interpret the coefficient on DrivingTime in your regression in question 7. Is the magnitude of the coefficient plausible?

10. Finally, we might expect the relationship between math scores and house prices to be different for large homes than small homes.

    a. Articulate why this might be the case.

    b. How might you use regression analysis to test the hypothesis that the relationship between math scores and house prices is different for large homes than small homes? What do you find?

11. Don't forget to close your log file! Go to "File → Log → Close."