Econometrics
Homework 1
DUE: March 4th

You are encouraged to work with your classmates on this homework. However, you each must turn in your own answers. I will not accept identical answers from multiple students nor do I want to see the identical answers from the back of the text.

Part I – A few problems to review statistics
1. Chapter 16, number 2
2. Chapter 16, number 8
3. Chapter 16, number 10
4. Chapter 16, number 13

Part II – Questions from the text
1. Chapter 1, number 4
2. Chapter 1, number 7
3. Chapter 1, number 8
4. Chapter 2, number 4
5. Chapter 2, number 6
6. Chapter 2, number 9
7. Chapter 2, number 11
8. Chapter 3, number 3
9. Chapter 3, number 5
10. Chapter 3, number 10
11. Chapter 4, number 4
12. Chapter 4, number 5
13. Chapter 4, number 9
14. Chapter 5, number 2
15. Chapter 5, number 4
16. Chapter 5, number 8
17. Chapter 5, number 10
18. Chapter 5, number 12
19. Chapter 5, number 15
Part III – Using data – Male Fruitflies

This dataset is called ‘fruitfly.dta’. It is already in Stata format. You must use Stata for the following questions. Here is a list of the variables in the dataset and their descriptions:

- **ID**: This is just an identification number from 1 – 125
- **PARTNERS**: This is the number of sexual companions that the fruitfly had (0, 1, or 8)
- **TYPE**: This is the type of companion (0=newly pregnant female, 1=virgin female, 9=not applicable, i.e. PARTNERS=0)
- **LNGEVITY**: This is lifespan, in days
- **THORAX**: Length of the thorax, in mm
- **SLEEP**: Percentage of each day spent sleeping

1. I am listing this first as a friendly reminder. You must create and print out the stata log file. Edit it where appropriate.

2. Examine the data. Briefly discuss how the data is set up.

3. Create a dummy variable called ‘ACTIVE’ that is =1 if the fruitfly has at least one partner and =0 if he has no sexual partners. Create a dummy called ‘ACTIVE1’ if PARTNERS=1. Create a dummy called ‘ACTIVE8’ if PARTNERS=8. Why do we not need to create a variable called ‘ACTIVE0’?

4. Report the means, standard deviations, and medians of all variables. Discuss the meaning of the mean of the TYPE, PARTNERS, and ACTIVE variables.

5. Report and compare the means, standard deviations, and medians for the LNGEVITY variable if the fruitfly has 0, 1, and 8 partners. Does this suggest a relationship between sexual behavior and lifespan?

6. Suppose you are interested in finding the determinants of lifespan for male fruitflies. You will test LNGEVITY = f(PARTNERS, THORAX, SLEEP). What do you predict for the sign of the coefficients on the independent variables?

7. Run an OLS regression with LNGEVITY as your dependent variable and ACTIVE1, ACTIVE8, THORAX, and SLEEP as your independent variables and report the coefficients and t-stats for these variables as well as the constant in a table. Also report the F-stat and adjusted r-squared. Discuss the results. Were your predictions correct? What is the meaning of the coefficients? What variables are significant at the .05 level?

8. Run another OLS regression with LNGEVITY as your dependent variable and THORAX, SLEEP, and TYPE as your dependent variables for those you are
active only. Prepare another table similar to the table needed for question 7 above. Discuss these results.

9. What conclusions can you make about the sexual and sleeping behavior of male fruitflies and their lifespan?

Part IV – Using data – Cars

This dataset is called ‘93cars.dta’. It is already in Stata format. You must use Stata for the following questions. Here is a list of the variables in the dataset and their descriptions:

MANUFACT  Manufacturer
MODEL  Model
TYPE  Small, Sporty, Compact, Midsize, Large
PRICE  Average in thousands of $
CITY_MPG  Miles per gallon, city
HIGH_MPG  Miles per gallon, highway
AIRBAGS  Airbags? (0=none, 1=driver only, 2=driver & passenger)
DRIVE_TR  Drive train (0=rear wheel drive, 1=front wheel drive, 2=all wheel drive)
CYLINDER  Number of cylinders
ENGSIZE  Engine size (liters)
HORSEPOW  Horsepower
RPM  Revolutions per minute at maximum horsepower
MANUAL  Manual transmission available?
FUELTANK  Fuel tank capacity in gallons
PASSENGER  Passenger capacity
LENGTH  Length in inches
WHEELBASE  Wheelbase in inches
WIDTH  Width in inches
U_TURN  U-turn space in feet
REARSEAT  Rear seat room in inches
LUGGAGE  Luggage capacity in cubic feet
WEIGHT  Weight in pounds
DOMESTIC  Domestic?

1. I am listing this first as a friendly reminder. You must create and print out the stata log file. Edit it where appropriate.

2. Examine the data. Briefly discuss how the data is set up. In particular, look at the price data. Are there any outliers in the data? In other words, are there any models that are considerably higher or considerably lower than the remainder of the data points?
3. Create Dummies for each TYPE. Create a dummy called ‘AIRBAG’ =1 if any airbags. Create dummies for each different CYLINDER possibility. After you complete this step, drop the TYPE, AIRBAGS, and CYLINDER variables.

4. Report the means, standard deviations, and medians of all variables.

5. Report and compare the means, standard deviations, and medians for PRICE, CITY_MPG, and HIGH_MPG for each type. Discuss your results.

6. Report and compare the means, standard deviations, and medians for PRICE, CITY_MPG, and HIGH_MPG for domestic vs. foreign cars. Discuss your results. Run hypothesis tests to determine whether there are significant differences in these variables between the 2 groups.

7. Report the correlation coefficients for the variables. Discuss your findings.

8. Suppose you are interested in what determines city and highway mpg. Discuss which variables should you include on the right hand side of these regressions. Be careful to discuss the concept of correlation vs. causation here.

9. Run an OLS regression with CITY_MPG as your dependent variable and any appropriate independent variables of your choosing. Report the appropriate coefficients, t-stats, adjusted r-squared, and f-stat in a table. Discuss the results.

10. Repeat question 8 for HIGH_MPG. Compare your results to those of question 8.

11. Now suppose you are interested in the determinants of price. Predict which variables should matter and their coefficients’ expected signs.

12. Run an OLS regression with PRICE as your dependent variable and the type variables, HIGH_MPG, AIRBAG, the cylinder variables, HORSEPOW, PASSENGE, FUELTANK, U_TURN, LUGGAGE, and DOMESTIC as independent variables. Report everything appropriate and discuss the results.

13. Run another OLS regression on PRICE as in question 12 but eliminating independent variables with insignificant coefficients. Report and discuss. Is this regression more reliable than question 12’s?

14. Repeat question 13 dropping any outliers you may have encountered. Are there any major differences between these results and those from question 13?

15. What conclusions can you make about the determinants of car price and gas efficiency?