Ecn 102 - Analysis of Economic Data
University of California - Davis
John Parman

January 22, 2011

Problem Set 3

This problem set will be due by 5pm on Monday, February 7th. It may be turned in either in class or put in your TA's mailbox in the economics department mailroom. You may work in groups but everyone in the group must write up their own solutions including creating their own graphs and tables. Please only turn in what the questions asks for (do not print out and turn in the tables of raw data or variables you construct).

---

1. **Fuel Economy and Car Size** For this question, use the data on fuel economy for 2009 model cars. These data are contained in the file '2009-fuel-economy.csv' on Smartsite in the data folder. Descriptions of the variables are given in the file 'fuel-economy-variables.csv'.

   (a) The EPA's fuel economy data distinguishes between two-door, four-door and hatchback vehicles when specifying passenger and cargo space. As a result, in the data's current form, different variables capture car size for different car types. For our analysis, we will focus on four-door vehicles. Sort the data by the four-door passenger space variable and keep only those observations with a positive value reported for this variable (deleting the observations for two-door cars and hatchbacks). Create a new variable the measures that combined passenger and cargo space for the four-door cars.

   (b) Create a histogram and a table of descriptive statistics for highway fuel efficiency (highway miles per gallon). What is the mean miles per gallon for 2009 four-door vehicles?

   (c) Create a scatterplot showing the relationship between highway fuel efficiency and car size. Fuel efficiency should be measured on the vertical axis in terms of highway miles per gallon. Car size should be measured by the combined passenger and cargo space.

   (d) Run a regression of fuel efficiency on car size. Highway miles per gallon should be your dependent variable and total volume (passenger plus cargo space) should be your independent variable. Interpret your regression results. Does the sign of your coefficient for car size make sense? Would you expect this to be a causal relationship (changes in car size cause changes in fuel efficiency)? What other variables may be influencing the relationship? That is to say, what variables have we not included in our regression that would be correlated with both car size and fuel efficiency?

   (e) Repeat part (d) using engine size rather than car volume as the measure of car size. Engine size is given by the displacement variable and is measured in liters.

(f) Which measure of car size seems to be a better predictor of fuel efficiency? Explain your answer.

2. **Fuel Economy Over Time** For this question, use the data on fuel economy for both 2009 model cars and for 2001 model cars. These data are contained in the files '2009-fuel-economy.csv' and '2001-fuel-economy.csv', respectively. The variables are the same in the two files.

   (a) We want to look at fuel economy over time. To do this, we'll combine the 2001 and 2009 data into a single dataset. Before we combine the data, we need to construct a new variable in each dataset giving the year. Create a variable named YEAR in each dataset, taking on the value 2001 in the 2001 data and the value 2009 in the 2009 data. Now copy and paste the data from one file into the other so that you have a single dataset with both the 2001 and 2009 data (make certain you line up the variables correctly when pasting the data).

   (b) For this entire question, we want to focus on a handful of car types and manufacturers and see how fuel economy has changed over time. Sort the data by class and keep the observations for compact, midsize and large cars, deleting all other observations. Now sort the data by manufacturer and keep only the observations for Ford, Chevrolet, Honda, Toyota, BMW and Mercedes. This will leave us with a much more manageable size for the dataset.

   (c) First we want to see if fuel economy has improved between 2001 and 2009. Create a table showing the average miles per gallon for city driving in 2001 and in 2009, the average miles per gallon for highway driving in 2001 and in 2009 and the average miles per gallon for combined city and highway driving in 2001 and in 2009 for each car type. Your table should have 18 different miles per gallon numbers in it (3 mpg measures x 3 car types x 2 years). Based on this table, does it look like fuel economy has improved over the past decade?

   (d) Create a scatterplot showing combined city and highway miles per gallon (the COMB MPG (GUIDE) variable) on the vertical axis and engine size, measured by displacement in liters on the horizontal axis. Include two different sets of points on your graph, one set for the year 2001 observations and one set for the year 2009 observations. Based on the scatterplot, does it appear that fuel economy has improved over the past decade?

   (e) Run a regression of fuel economy on engine size, using combined city and highway miles per gallon as your dependent variable and displacement as your independent variable using the year 2001 observations. Run the same regression again but use the year 2009 observations. Based on your regression results, explain whether there is evidence of improvements or declines in fuel economy over the past decade. Consider both the coefficient on displacement and the value for the intercept when coming up with your answer.

(f) Create a table showing mean displacement by car type for 2001 and 2009. Explain how the values in this table help reconcile differences in your answers for parts (c) and (e).

(g) The Environmental Protection Agency changed their methods of calculating gas mileage in 2008. According to the EPA, they changed the way they computed gas mileage to take into account the following:

- Faster speeds and acceleration
- Air conditioner use
- Colder outside temperatures

How does this information change your evaluation of how fuel economy has changed between 2001 and 2009?