

## **Econ 400: Stata Assignment 2**

Due July 20, 12:00 pm. Please turn in a hard copy (i.e. not an e-mail).

For this assignment you will use the data set “house.dta”. The data are a random sample of records of resales of homes from Feb 15 to Apr 30, 1993 from the files maintained by the Albuquerque Board of Realtors. This type of data is collected by multiple listing agencies in many cities and is used by realtors as an information base. (I found the data at <http://lib.stat.cmu.edu/DASL/Datafiles/homedat.html>)

Complete the tasks listed below. Use a Stata log file to record your results, and print out all STATA output. Feel free to write out the answers to the short answer questions on the Stata Log printout if you can do so legibly; otherwise you can cut and paste the log file into a Word document and type out your answers where appropriate.

### **I. Descriptive Statistics**

- (1) Use the ‘describe’ command to familiarize yourself with the data set
- (2) Get summary statistics for all variables in the data set. What is the sample mean for price? Sample standard deviation?
- (3) Create a table that presents number of observations, sample mean and sample standard deviation for price, and the sample mean and sample standard deviation of square footage (‘sqft’) separately for houses in the NE and not in the NE.
  - a. How does price and square footage compare for houses in the NE vs houses not in the NE?

### **II. Estimation & Hypothesis Testing**

- (1) Estimate a 95% confidence interval for the average price of a house in Albuquerque.
  - a. What is the point estimate?
  - b. What is the standard error for the sampling distribution of the sample mean?
  - c. What are the upper and lower bounds of the 95% confidence interval?
- (2) Estimate a 99% confidence interval for average price of a house. How does this interval compare to the 95% confidence interval?
- (3) Estimate a 95% confidence interval for average price of a house in the NE quadrant. Estimate a 95% confidence interval for population mean earnings for a house that is not in the NE quadrant.
  - a. How do the interval estimates compare?

- (4) Test the hypothesis that the mean price of a house in Albuquerque is greater than \$100,000.
- What is the null hypothesis? Alternative Hypothesis?
  - What is the test statistic?
  - What is the p-value of this test?
  - Does the evidence suggest that you can reject the null hypothesis at the .05 level of significance? At the .01 level?
- (5) Test the hypothesis that the mean price of a house in the NE quadrant of Albuquerque is larger than the mean price of a house in Albuquerque that is not in the NE quadrant.
- What is the null hypothesis? Alternative Hypothesis?
  - What is the test statistic?
  - What is the p-value of this test?
  - Does the evidence suggest that you can reject the null hypothesis at the .05 level of significance? At the .01 level?

### III. Regression

In this section you will examine the relationship between price of a house in Albuquerque and square footage.

- Is price and square footage of a house correlated? What is the correlation coefficient?
- Graph a scatter plot of price vs. square footage
- Regress price on square footage.
  - How many observations are used in the regression?
  - What is the R-squared? Adjusted R squared?
  - What is the estimated coefficient on square footage? Interpret this coefficient
  - Is the estimated coefficient on square footage statistically significant at the 5% level?
- Graph a scatter plot of price vs. square footage and include the linear regression line