

Homework 2

INSTRUCTIONS: The deadline for this assignment is **Wednesday, May 13** at 4:30 pm. Submit the log and output files with your assignment.

1. Question 3.4 on page 69 of the text.
2. Question 4.3 on page 97.
3. Question 5.3 on page 127. (Note: there are 2 parts marked (a). Answer both, as well as parts (b)-(c)). Also, interpret the value of R-squared verbally.
4. This question uses the data set, br2.dat, that was used in Homework 1. (See Homework 1 for data definitions.)
 - a. Using the data for all 1080 houses in the sample, estimate the following model:
$$\text{Price}_t = \beta_1 + \beta_2 \text{Age}_t + e_t$$
 - b. Interpret the estimated coefficient on Age.
 - c. Interpret the value of R-squared verbally.
 - d. Test the significance of the parameter estimate on Age using a 95 percent confidence interval.
 - e. Obtain confidence intervals using SAS. Does the confidence interval for the estimated parameter on Age from SAS match the one you estimated in part (d)?
 - f. Using SAS, obtain prediction intervals and plot price and age with the regression line and prediction interval bounds.
 - g. Estimate the multiple regression, $\text{Price}_t = \beta_1 + \beta_2 \text{Age}_t + \beta_3 \text{Sqft}_t + e_t$.
 - h. Interpret the coefficients on Age and Sqft in the multiple regression.

524 Students - Project Proposal

The project proposal is designed to keep you on track for your project. It will help us to see if your plan is reasonable and will help you to make progress. Hopefully, part or all of this proposal will be integrated into your final paper. The elements that you should include are:

- Statement of the primary **question** that you will address
- Brief statement of your **Motivation** for the topic, e.g.,
why it is important; who might be interested in the results; policy implications; *or* why you chose the topic. (Because we only have 10 weeks to do the project, it is acceptable in this class to use “data availability” as the reason that you chose the topic.)
- The **regression equation** that you will estimate to analyze the question. Write it out in equation form. Replace X and Y with the abbreviations for the variables that you will use.

- State the full name and reference for the **data** source. Define the variables in the regression equation in detail, including units of measure and how the variables are constructed. Produce the means, standard errors, mins and maxs, for the variables that you are considering for your project. If you have categorical data with a limited range of values, print frequency tables:
PROC FREQ;
Tables X1 X2; (X1 and X2 are the variable names- you can also add other variables)
run;
Make sure that the values of the statistics make sense, e.g., in some data sets missing values are coded -999 which will be read as legitimate values when you run your regression—checking the mean and min will reveal this. Another case— a dummy variable with a max of 1000.
- **Expected results**, focusing on the implications for the primary question. Referring to the regression equation, indicate what signs you expect to find on the coefficients when you estimate the regression model.

Upcoming Project Assignment for Homework 3

This assignment may also be integrated into your final project. Please submit the following with Assignment 3 or earlier:

- A draft of the **data section** of your paper. It should include information on the data from Homework 2, with a table of descriptive statistics: means and standard deviations for continuous variables. For discrete variables (i.e., dummy variables or variables that take on whole number values, e.g., 1,2,3,4,5) list frequencies (i.e., counts/number of observations for the variable) for each value of the variable, and list number of observations for the variable, instead of means and standard deviations.
- **Preliminary results**, and a discussion of the results including surprises regarding your expected results, if any, and what the preliminary results indicate about your research question.