EC 339

Problem Set 1

Prof. Santetti

Fall 2022

INSTRUCTIONS: Carefully read all problems. You must submit a single R script (Section 001) and `Stata` do-file (Section 002) with your *first name* (mine would be `marcio.R` and `marcio.do`). In case you submit your files with different names, you will lose 1 point.

You can find templates for your answer scripts/do-files on `theSpring`, under the "Templates" module. Please consider using it.

I should be able to fully replicate your code to answer the questions, as well as fully understand your written interpretations to the proposed problems.

Avoid using unnecessary code in your submission files. It is totally fine to do other things by yourself that may help you better understand the data and the problems. However, for grading purposes, I am only interested in the commands and interpretations that actually answer the questions. You may keep a separate file for yourself with your additional explorations.

`Assignment due October 12 (W), before class`.
`Points Possible: 30`

- You have 2 weeks to complete this assignment. In accordance with our `course syllabi`, no late submissions will be accepted.
- Be honest. Don't cheat.
- As a Skidmore student, always recall your votes of academic integrity, and the **`Honor Code`** you have abided by:

> *"I hereby accept membership in the Skidmore College community and, with full realization of the responsibilities inherent in membership, do agree to adhere to honesty and integrity in all relationships, to be considerate of the rights of others, and to abide by the college regulations."*

Have fun!

Problem 1

Consider the following regression model:

$$\widehat{log(price_i)} = 4.39 + 0.036\ sqft_i$$

where $price_i$ is the selling price of a property (in $1,000), and $sqft_i$ is the total interior square footage of a house (in 100 sqft). This data set has 500 observations of single-family home sales in Baton Rouge, LA in 2013.

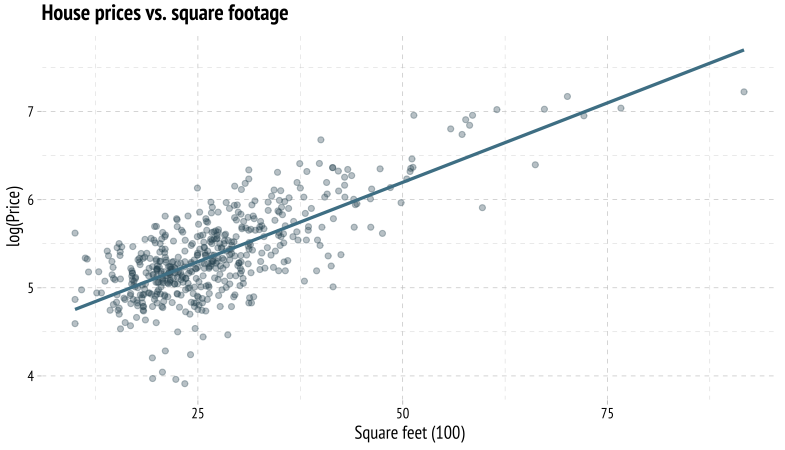
(a) Interpret this regression's *slope* coefficient. (*2 points*)

(b) How many *degrees-of-freedom* remain after this estimation? Explain. (*2 points*)

(c) What other variables would you consider to be included in this regression's *error term*? Explain. (*2 points*)

(d) This regression's *coefficient of determination* (R^2) is 0.5417. Interpret its value. (*2 points*)

(e) Based on the scatter plot below, do you believe this is an appropriate *functional form* to describe this relationship? Explain. (*2 points*)



Problem 2

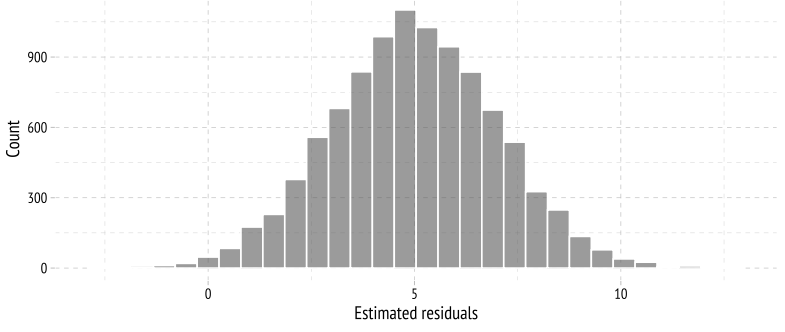
Consider the following model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

Answer the following questions:

(a) Is this model linear in parameters? Explain. (*2 points*)

(b) After being estimated, a histogram of the model's residuals looks like the one below:



Is everything looking as it should, given Classical Assumption II? Explain. (*2 points*)

(c) The relationship between variables x_2 and x_3 is described by:

$$x_{2i} = 5.57 - 2x_{3i} + \epsilon_i$$

where ϵ_i is a random variable with mean 2 and constant variance. With this information, does this model violate the *no perfect multicollinearity* assumption? Explain your answer. (*2 points*)

(d) Suppose a colleague sends you an email telling you that they have estimated the same model, but adding variable x_4 has increased the coefficient of determination (R^2) from .36 to .45. Their claim is that you should stick with this new model. Give at least two possible replies to your friend's argument. (*2 points*)

(e) This same colleague would like to estimate the elasticity of the dependent variable, *y*, with respect to x_1. They do not know how to proceed. In your own words, explain the procedure to your friend, showing *how* the model should be estimated, and *why* it should look the way is does. (*2 points*)

Problem 3

The `koop_tobias` data set brings a subset of the data used by `Koop and Tobias (2004)`. The sample is restricted to white males who are at least 16 years old and who worked at least 30 weeks and 800 hours during the year of 1987.[1] You can also find a `.txt` file describing the variables.

(a) After importing the data and getting some acquaintance with it, estimate the following regression model: (*2 points*)

$$log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 fatheduc_i + \beta_3 exper_i + u_i$$

(b) Interpret the above regression's slope coefficients. (*2 points*)

(c) Now, add a *proxy* variable for worker's *ability* to the previous model, in this data set denoted as *score*. This variable is constructed from the 10 component tests of the Armed Services Vocational Aptitude Battery, administered in 1980, and standardized for age. (*2 points*)

(d) From part (c)'s model, interpret the effect of *score* on the dependent variable. (*2 points*)

(e) Critically compare the *goodness-of-fit* measures between your models from parts (a) and (c). (*2 points*)

[1] This data set is part of Carter-Hill, Griffiths, and Lim, *Principles of Econometrics*, 2018, 5th edition, Wiley.