

Homework 4

Simple Regression Model (30 points)

Instruction:

- This HW must be done in Rmarkdown!
- Please submit both the .rmd and the Microsoft word files. (Do not submit a PDF or any other image files as the TAs are going to give you feedback in your word document)
- Name your files as: HW4-groupnumber-name
- All the HW assignments are individual work. However, I highly encourage you to discuss it with your group members.
- Late homework assignments will not be accepted under any circumstances.

Problems

Question 1 Using data from 1988 for houses sold in Andover, Massachusetts, from Kiel and McClain (1995), the following equation relates housing price (*price*) to the distance from a recently built garbage incinerator (*dist*):

$$\widehat{\log(\text{price})} = 9.40 + 0.132 \log(\text{dist})$$
$$n = 135, R^2 = 0.162$$

- (i) Interpret the coefficient on $\log(\text{dist})$. Is the sign of this estimate what you expect it to be?
- (ii) What other factors about a house affect its price? Might these be correlated with distance from the incinerator?
- (iii) Do you think simple regression provides an unbiased estimator of the ceteris paribus elasticity of price with respect to *dist*? (Think about the city's decision on where to put the incinerator. Hint: which of the SLR assumptions is/are violated?)

Question 2 Consider the savings function:

$$\text{sav} = \beta_0 + \beta_1 \text{inc} + u, \quad u = \sqrt{\text{inc}} \cdot e$$

where e is a random variable with $E(e) = 0$ and $\text{var}(e) = \sigma_e^2$. Assume that e is independent of *inc*.

- (i) Show that $E(u|\text{inc}) = 0$, so that the key zero condition mean assumption is satisfied.
- (ii) Show that $\text{Var}(u|\text{inc}) = \sigma_e^2 \text{inc}$, so that the homoskedasticity Assumption SLR.5 is violated. In particular, the variance of *sav* increases with *inc*. [Hint: $\text{Var}(e|\text{inc}) = \text{Var}(e)$ if e and *inc* are independent].
- (iii) Provide a discussion that supports the assumption that the variance of savings increases with family income.

Computer Exercises

Question 3 The data in 401K are a subset of data analyzed by Papke (1995) to study the relationship between participation in a 401(k) pension plan and the generosity of the plan. The variable *prate* is the percentage of eligible workers with an active account; this is the variable we would like to explain. The measure of generosity is the plan match rate, *mrte*. This variable gives the average amount the firm contributes to each worker's plan for each \$1 contribution by the worker. For example, if *mrte* = 0.5 then a \$1 contribution by the workers is matched by a \$0.50 contribution by the firm.

- (i) Find the average participation rate and the average match rate in the sample of plans.
- (ii) Now, estimate the simple regression equation

$$\widehat{prate} = \hat{\beta}_0 + \hat{\beta}_1 mrte$$

and report the results along with the sample size and R-squared

- (iii) Interpret the intercept in your equation. Interpret the coefficient on *mrte*.
- (iv) Find the predicted *prate* when *mrte* = 3.5. Is this a reasonable prediction? Explain what is happening here.
- (v) How much of the variation in *prate* is explained by *mrte*? Is this a lot in your opinion?

Question 4 Use the data in MEAP93 for this question. We want to explore the relationship between the math pass rate (*math10*) and spending per student (*expend*)

- (i) Do you think each additional dollar spent has the same effect on the pass rate, or does a diminishing effect seem more appropriate? Explain.
- (ii) In the population model

$$math10 = \beta_0 + \beta_1 \log(expend) + u$$

Argue that $\frac{\beta_1}{10}$ is the percentage point change in *math10* given a 10% increase in *expend*

- (iii) Use the data in MEAP93 to estimate the model from part (ii). Report the estimated equation in the usual way, including the sample size and R-squared
- (iv) How big is the estimated spending effect? Namely, if spending increases by 10%, what is the estimated percentage point increase in *math10*.
- (v) One might worry that regression analysis can produce fitted values for *math10* that are greater than 100. Why is this not much of a worry in this data set?

Question 5 Use the data in CHARITY [obtained from Franses and Paap (2001)] to answer the following questions:

- (i) What is the average gift in the sample of 4,268 people (in Dutch guilders)? What percentage of people gave no gift?
- (ii) What is the average mailings per year? What are the minimum and maximum values?
- (iii) Estimate the model

$$gift = \beta_0 + \beta_1 mailsyear + u$$

by OLS and report the results in the usual way, including the sample size and R-squared.

- (iv) Interpret the slope coefficient. If each mailing costs one guilder, is the charity expected to make a net gain on each mailing? Does this mean the charity makes a net gain on every mailing? Explain.

Question 6 The data set in CATHOLIC includes test score information on over 7,000 students in the United States who were in eighth grade in 1988. The variables *math12* and *read12* are scores on twelfth grade standardized math and reading tests, respectively.

- (i) How many students are in the sample? Find the means and standard deviations of *math12* and *read12*
- (ii) Run the simple regression of *math12* on *read12* to obtain the OLS intercept and slope estimates. Report results in the form:

$$\widehat{math12} = \hat{\beta}_0 + \hat{\beta}_1 read12$$

$$n = ?, R^2 = ?$$

where you fill in the values for $\hat{\beta}_0$ and $\hat{\beta}_1$ and also replace the question marks.

- (iii) Does the intercept reported in part (ii) have a meaningful interpretation? Explain.
- (iv) Are you surprised by the $\hat{\beta}_1$ you found? What about R^2 ?
- (v) Suppose that you present your findings to a superintendent of a school district, and the superindtendent says, "Your findings show that to improve math scores we just need to improve reading scores, so we should hire more reading tutors." How would you respond to this comment? (Hint: If you instead run the regression of *read12* on *math12*, what would you expect to find?)