

ECON 3161 Homework 3 (Fall 2023)

Due Date: Oct 15th (Sunday)

Instruction: There are 80 points in total. You can either type or handwrite your answer. Then upload your answer file on canvas. The question 2 is a coding problem. You could use any software you feel comfortable with. However, Excel is not recommended. Attach (or copy) your code to the end of your answer file.

1. (20 pts) Using 722 household information, we estimate the following equation:

$$\widehat{educ} = 10.36 - 0.094sibs + 0.131meduc + 0.210feduc, \quad (1)$$

where *educ* is years of schooling, *sibs* is the number of siblings, *meduc* is the mother's years of schooling, and *feduc* is father's years of schooling. Answer the following questions:

(1) Does *sibs* have the expected effect? Explain. Holding *meduc* and *feduc* fixed, by how much does *sibs* have to increase to reduce predicted years of education by one year? (An approximate number is okay here) (5pts)

(2) Discuss the interpretation of the coefficient on *meduc*. (5pts)

(3) Suppose that man A has no siblings, and his mother and father each have 12 years of education, and man B has no siblings, and his mother and father each have 16 years of education. What is the predicted difference in years of education between the two men? (5pts)

(4) What are other factors might affect a person's education? Are they likely to be correlated with the three regressors in equation (1)? (5 pts)

2. (30 pts) The following question is an computer exercise. You're expected to use the software to estimate the following model:

$$price = \beta_0 + \beta_1sq\,ft + \beta_2bdrms + u, \quad (2)$$

where *price* is the house price measured in thousands of dollars, *sqft* is the square feet of the house, and *bdrms* is the number of bedrooms. The dataset HPRICE is uploaded on canvas (I provide both the .dta and .xlsx versions).

(1) Estimate model (2) using the OLS method (hint: for stata software, you will use the `reg` command; for R software, you can use `lm` command) . Write out the results in equation form. (6 pts)

Note: Your answer should include the estimated coefficients as well as the standard errors. You can put your standard errors in a parenthesis below the corresponding coefficients. I give the example here:

$$\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 sqft + \hat{\beta}_2 bdrms$$

$$(se(\hat{\beta}_0)) \quad (se(\hat{\beta}_1)) \quad (se(\hat{\beta}_2))$$

(2) What is the estimated increase in price for a house with one more bedroom, holding square footage constant? (6 pts)

(3) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? (6 pts)

(4) The first house in the sample has $sqft = 2,438$ and $bdrms = 4$. Find the predicted selling price for this house. (6 pts)

(5) The actual selling price of the first house in the sample was \$300,000 (so $price = 300$). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house? (6 pts)

3. (15 pts) In a study relating college grade to time spent in various activities, you distribute a survey to several students. The students are asked how many hours they spend each week in four activities: studying, sleeping, working, and leisure. Any activity is put into one of the four categories, so that

(1) In the model

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + \beta_4 leisure + u, \quad (3)$$

does it make sense to hold *sleep*, *work*, *leisure* fixed, while changing *study*? (5 pts)

(2) Explain why this model violates the perfect collinearity? (5 pts)

(3) How would you reformulate the model so that its parameters have a useful interpretation and it satisfies no perfect collinearity assumption? (5 pts)

4. (15 pts) What factors affect the variance of the OLS estimators in a multiple regression? List these factors. Furthermore, explain how they affect the variance.