ECON 3161 Fall 2023

Homework 3

Vidit D. Pokharna Due: October 15th 2023

- **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 points) Using 722 houshold information, we estimate the following equation:

$$\hat{educ} = 10.36 - 0.094sibs + 0.131meduc + 0.210feduc,$$

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:

(i) 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)

Solution:

The coefficient of *sibs* is -0.094. A negative coefficient suggests that as the number of siblings increases, the predicted years of education decrease. This effect is expected and consistent with common intuition: a larger number of siblings may result in less individual attention and resources available for each child, which can potentially lead to fewer years of education.

Reducing predicted years of education by one year, you would need to increase *sibs* by $(-1)/(-0.094) \approx \boxed{10.64}$.

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

Solution:

$(1)/(0.131) \approx 7.63$.

Thus, for every 7.63 more years a mother indulges in schooling, the years of schooling would increase by a year, holding *sibs* and $feduc$ fixed.

(iii) 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)

Solution:

$(0.131 + 0.210)(16 - 12) = \boxed{1.364}$

(iv) 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)

Solution:

There are factors that can be correlated with the current regressors in the equation, such as parental socioeconomical status (dependent on parents' schooling), and certain factors that have no correlation with the current regressors, such as peer influence on the student.

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

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

where *price* is the house price measured in thousands of dollars, $sqrft$ 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).

(i.) 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)

Solution:

$price = -19.315 + .1284362(sqrft) + 15.19819(bdrms) + u$

```
. use "C:\Users\vpokharna3\OneDrive - Georgia Institute of Technology\HPRICE.dta
> "

. reg price sqrft bdrms
```

Source	SS	df	MS	Number of obs	=	88
				F(2, 85)	=	72.96
Model	580009.152	2	290004.576	Prob > F	=	0.0000
Residual	337845.354	85	3974.65122	R-squared	=	0.6319
				Adj R-squared	=	0.6233
Total	917854.506	87	10550.0518	Root MSE	=	63.045

price	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
sqrft	.1284362	.0138245	9.29	0.000	.1009495	.1559229
bdrms	15.19819	9.483517	1.60	0.113	-3.657582	34.05396
_cons	-19.315	31.04662	-0.62	0.536	-81.04399	42.414

```
.
```

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

Solution: This would be the coefficient of *bdrms*, which is ≈ 15.20, or $\boxed{\$15200}$

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

Solution:

$.1284362(140) + 15.19819(1) = 33.179258 \approx 33.18$ or $\boxed{\$33180}$

(iv.) The first house in the sample has $sqrft = 2438$ and $bdrms = 4$. Find the predicted selling price for this house. (6 pts)

Solution:

Predicted $= 0.1284362(2438) + 15.19819(4) - 19.315 =$ $\boxed{\$354.6052156}$

(v.) The actual selling price of the first house in the sample was \$300000 (so $price = 300$). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house? (6 pts)

Solution:

Actual = \$300

Predicted = \$354.6052156

Residual $= 1000(300 - 354.6052156) =$ $\boxed{-\$54605.2156}$

This suggests that the buyer $\boxed{underpaid}$ for the house, since the residual is negative.

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

(i.) In the model

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

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

Solution:

It does not make sense to hold these constant because they are likely to have inter-dependencies or associations with the GPA. Additionally, with changes in studying time, there will likely be reverse changes in the other three activities (less time for the other activities). Thus, it does not make sense to hold sleep, work and leisure fixed for most practical scenarios.

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

Solution:

In this model, we have GPA as the dependent variable and four regressor variables: study, sleep, work, and leisure. The problem arises because the sum of the hours spent in these four activities is limited to 168 hours in a week (assuming a standard week with 24 hours per day). If we know the values of three of these variables, we can predict the fourth one by subtracting the sum of the three known variables from 168. For example, if we know how much time a student spends on studying, sleeping, and working, we can easily calculate the time spent on leisure. Therefore, a linear combination can be determined and the model will violate no perfect collinearity.

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

Solution:

We can reformulate this to combine alike variables. For example, we can combine work and study into productive time and sleep and leisure into leisure time. This will lower the chance of dependency and linearity between the variables and help satisfy the no perfect collinearity assumption.

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.

Solution:

1. Sample size: As the sample size increases, the variance of the OLS estimators tends to decrease.
2. Number of independent variables: The variance of the OLS estimators tends to increase as the number of predictors (independent variables) in the model increases.
3. Outliers: Outliers can have a significant impact on the OLS estimators. They can increase the variance of the estimators by pulling the estimates towards the outliers, making the coefficients less representative of the majority of the data.