Student: ____________________ **Date:** ____________________	**Instructor:** Richeng Piao **Course:** ECON 2560 - Applied Econometrics	**Assignment:** Practice Problem Set 6

1. A researcher is interested in the effect on test scores of computer usage. Using school district data, she regresses district average test scores on the number of computers per student.

   What are possible sources of bias for $\widehat{\beta}_1$, the estimated effect on tests scores of increasing the number of computers per student? For each source of bias below, determine whether $\widehat{\beta}_1$ will be biased up or down.

   1. Average income per capita in the district. If this variable is omitted, it will likely produce a(an) (1) ________ bias of the estimated effect on tests scores of increasing the number of computers per student.

   2. The availability of computerized adaptive learning tools in the district. If this variable is omitted, it will likely produce a(an) (2) ________ bias of the estimated effect on tests scores of increasing the number of computers per student.

   3. The availability of computer-related leisure activities in the district. If this variable is omitted, it will likely produce a(an) (3) ________ bias of the estimated effect on tests scores of increasing the number of computers per student.

   (1) ○ upward ○ downward
   (2) ○ upward ○ downward
   (3) ○ upward ○ downward

   Answers (1) upward
   (2) upward
   (3) downward

   ID: Review Concept 6.1

2. A researcher plans to study the causal effect of police crime using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the country's police force.

Why is this regression likely to suffer from omitted variable bias?

- ○ **A.** There are other important determinants of a country's crime rate, including demographic characteristics of the population, that if left out of the regression would bias the estimated partial effect of the (per capita) size of the county's police force.
- ○ **B.** Other regressors are likely perfectly collinear with the (per capita) size of the county's police force, so they should be included in the regression.
- ○ **C.** There are other important determinants of a country's crime rate, including demographic characteristics of the population, that if left out of the regression would increase the variance of the estimated coefficient for the (per capita) size of the county's police force.
- ○ **D.** None of the above are correct.

Which of the following variables are likely useful to add to the regression to control for important omitted variables?
(*Check all that apply*)

- ☐ **A.** The number of bowling alleys in the county.
- ☐ **B.** The average level of education in the county.
- ☐ **C.** The fraction of young males in the county population.
- ☐ **D.** The average income per capita of the county.

Suppose that crime rate is positively affected by the fraction of young males in the population, and that counties with high crime rates tend to hire more police. Use the following expression for omitted variable bias to determine whether the regression will likely over– or underestimate the effect of police on the crime rate.

$$\hat{\beta}_1 \to_p \beta_1 + \rho_{Xu}\frac{\sigma_u}{\sigma_X}$$

- ○ **A.** The regression will likely overestimate β_1. That is, $\widehat{\beta_1}$ is likely to be larger than β_1.
- ○ **B.** The regression will likely underestimate β_1. That is, $\widehat{\beta_1}$ is likely to be smaller than β_1.

Answers A.

There are other important determinants of a country's crime rate, including demographic characteristics of the population, that if left out of the regression would bias the estimated partial effect of the (per capita) size of the county's police force.

B. The average level of education in the county., C. The fraction of young males in the county population., D. The average income per capita of the county.

A. The regression will likely overestimate β_1. That is, $\widehat{\beta_1}$ is likely to be larger than β_1.

ID: Exercise 6.6

3. Critique each of the following proposed research plans. Your critique should explain any problems with the proposed research and describe how the research plan might be improved. Include any discussion of any additional data that need to be collected.

A researcher is interested in determining whether a large aerospace firm is guilty of gender bias in setting wages. To determine potential bias, the researcher collects salary and gender information for all the firm's engineers. The researcher then plans to conduct a "difference in means" test to determine whether the average salary for women is significantly less than the average salary for men.

Which of the following variables are likely useful to add to the regression to control for important omitted variables? (*Check all that apply*)

- ☐ **A.** Work experience.
- ☐ **B.** Type of engineer.
- ☐ **C.** Education level.
- ☐ **D.** Eye color of the engineer.

A researcher is interested in determining whether the time spent in prison has a permanent effect on a person's wage rate. He collects data on a random sample of people who have been out of prison for at least 15 years. He collects similar data on a random sample of people who have never served time in prison. The data set includes information on each person's current wage, education, age, ethnicity, gender, tenure (time in current job), occupation, and union status, as well as whether the person was ever incarcerated. The researcher plans to estimate the effect of incarceration on wages by regressing wages on an indicator variable for incarceration, including in the regression other potential determinants of wages (education, tenure, union status, and so on).

Which of the following variables are likely useful to add to the regression to control for important omitted variables? (*Check all that apply*)

- ☐ **A.** Individual hair color.
- ☐ **B.** Gang activity.
- ☐ **C.** Excessive drug or alcohol use.
- ☐ **D.** Individual height.

Answers A. Work experience. , B. Type of engineer., C. Education level.

B. Gang activity. , C. Excessive drug or alcohol use.

ID: Exercise 6.7

---

4. A recent study found that the death rate for people who sleep 6 to 7 hours per night is lower than the death rate for people who sleep 8 or more hours. The 1.1 million observations used for this study came from a random survey of Americans aged 30 to 102. Each survey respondent was tracked for 4 years. The death rate for people sleeping 7 hours was calculated as the ratio of the number of deaths over the span of the study among people sleeping 7 hours to the total number of survey respondents who slept 7 hours. This calculation was then repeated for people sleeping 6 hours, and so on. Based on this summary, would you recommend that Americans who sleep 9 hours per night consider reducing their sleep to 6 or 7 hours if they want to prolong their lives? Why or why not? Explain.

Which of the following variables are likely useful to add to the regression to control for important omitted variables? (*Check all that apply*)

- ☐ **A.** Indicator for chronic illness.
- ☐ **B.** Individual eye color.
- ☐ **C.** Type of employment.
- ☐ **D.** Drug or alcohol use.

Answer: A. Indicator for chronic illness., C. Type of employment. , D. Drug or alcohol use.

ID: Exercise 6.8

---

5. Consider the multiple regression model with two regressors X_1 and X_2, where both variables are determinants of the dependent variable. You first regress Y on X_1 only and find no relationship. However, when regressing Y on X_1 and X_2, the slope coefficient $\hat{\beta}_1$ changes by a large amount.

This suggests that your first regression suffers from:

- **A.** omitted variable bias.
- **B.** heteroskedasticity.
- **C.** dummy variable trap.
- **D.** perfect multicollinearity.

Answer: A. omitted variable bias.

ID: Test A Ex 6.1.1

---

6. If you had a two-regressor regression model, then omitting one variable that is relevant:

- **A.** will always bias the coefficient of the included variable upward.
- **B.** can result in a negative value for the coefficient of the included variable, even though the coefficient will have a significant positive effect on Y if the omitted variable were included.
- **C.** makes the sum of the product between the included variable and the residuals different from 0.
- **D.** will have no effect on the coefficient of the included variable if the correlation between the excluded and the included variable is negative.

Answer: B.

can result in a negative value for the coefficient of the included variable, even though the coefficient will have a significant positive effect on Y if the omitted variable were included.

ID: Test B Ex 6.1.1

---

7. Consider the multiple regression model with two regressors X_1 and X_2, where both variables are determinants of the dependent variable.

When omitting X_2 from the regression, there will be omitted variable bias for $\hat{\beta}_1$:

- **A.** always.
- **B.** if X_1 and X_2 are correlated.
- **C.** if X_2 is measured in percentages.
- **D.** only if X_2 is a dummy variable.

Answer: B. if X_1 and X_2 are correlated.

ID: Test B Ex 6.1.2

---

8. An independent researcher is interested in finding out whether there exists a positive relationship between the number of years of formal education received by an individual and the number of years of formal education received by each of his parents. It is assumed that the number of years of formal education received by one parent of an individual is positively correlated with that of the other parent. The researcher randomly selects 180 individuals and estimates the following regression function:

$$\hat{Y}_i = 6 + 0.56X_i,$$

where Y_i denotes the number of years of formal education received by the i^{th} individual and X_i denotes the number of years of formal education received by the i^{th} individual's father.

Since the researcher only incorporates the educational attainment of an individual's father in the regression, and not that of the individual's mother, omitted variable bias will occur.

Which of the following statements correctly describes the omitted variable bias?

- **A.** Omitted variable bias arises when the variance of the conditional distribution of error term is not constant and depends on the omitted variable.
- **B.** Omitted variable bias arises when the omitted variable is correlated with a regressor and is a determinant of the dependent variable.
- **C.** Omitted variable bias arises when the omitted variable is a determinant of the independent variable but not of the dependent variable.
- **D.** Omitted variable bias arises when the omitted variable is correlated with the error term and is a determinant of a regressor.

Suppose the researcher somehow discovers that the values of the population slope (β_1), the standard deviation of the regressor (σ_X), the standard deviation of the error term (σ_u), and the correlation between the error term and the regressor (ρ_{Xu}) are 0.48, 0.61, 0.34, 0.47, respectively.

As the sample size increases, the value to which the slope estimator will converge to with high probability is ______.

(*Round your answer to two decimal places.*)

In this case, the direction of the omitted variable bias is (1) ______.

Assume a father's weight is correlated with his years of eduction, but is not a determinant of the child's years of formal education.

Which of the following statements describes the consequences of omitting the father's weight from the above regression?

- **A.** It will not result in omitted variable bias because the omitted variable, weight, is uncorrelated with the regressor.
- **B.** It will result in omitted variable bias the father's weight is a determinant of the dependent variable.
- **C.** It will not result in omitted variable bias because the omitted variable, weight, is not a determinant of the dependent variable.
- **D.** It will result in omitted variable bias because the omitted variable, weight, is correlated with the father's years of education.

(1) ○ positive
○ negative

Answers B.

Omitted variable bias arises when the omitted variable is correlated with a regressor and is a determinant of the dependent variable.

0.74

(1) positive

C. It will not result in omitted variable bias because the omitted variable, weight, is not a determinant of the dependent variable.

ID: Concept Exercise 6.1.1

---

9. A multiple regression includes two regressors:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

Use the tool palette to the right to answer the following questions.

What is the expected change in Y if X_1 increases by 6 units and X_2 is unchanged?

The expected change in Y if X_1 increases by 6 units and X_2 is unchanged is ______. *(Properly format your expression using the tools in the palette. Hover over tools to see keyboard shortcuts. E.g., a subscript can be created with the _ character.)*

What is the expected change in Y if X_2 decreases by 5 units and X_1 is unchanged?

The expected change in Y if X_2 decreases by 5 units and X_1 is unchanged is ______. *(Properly format your expression using the tools in the palette.)*

What is the expected change in Y if X_1 increases by 3 units and X_2 decreases by 10 units?

The expected change in Y if X_1 increases by 3 units and X_2 decreases by 10 units is ______. *(Properly format your expression using the tools in the palette.)*

Answers $6\beta_1$

$-5\beta_2$

$3\beta_1 - 10\beta_2$

ID: Review Concept 6.2

---

10. The data set consists of information on 4800 full-time full-year workers. The highest educational achievement for each worker was either a high school diploma or a bachelor's degree. The worker's ages ranged from 25 to 45 years. The data set also contained information on the region of the country where the person lived, marital status, and number of children. For the purposes of these exercises, let

AHE = average hourly earnings (in 2005 dollars)
College = binary variable (1 if college, 0 if high school)
Female = binary variable (1 if female, 0 if male)
Age = age (in years)
Ntheast = binary variable (1 if Region = Northeast, 0 otherwise)
Midwest = binary variable (1 if Region = Midwest, 0 otherwise)
South = binary variable (1 if Region = South, 0 otherwise)
West = binary variable (1 if Region = West, 0 otherwise)

Results of Regressions of Average Hourly Earnings on Gender and Education Binary Variables and Other Characteristics Using Data from the Current Population Survey

Dependent variable: average hourly earnings (*AHE*).

Regressor	(1)	(2)	(3)
College (X_1)	5.95	5.97	5.93
Female (X_2)	−2.88	−2.86	−2.86
Age (X_3)		0.32	0.32
Northeast (X_4)			0.75
Midwest (X_5)			0.65
South (X_6)			−0.29
Intercept	13.83	4.80	4.09
Summary Statistics			
SER	6.83	6.78	6.77
R^2	0.192	0.207	0.211
$\bar{R}^2$	0.192	0.207	0.210
n	4800	4800	4800

Using the regression results in column (2):

On average, a worker earns $ ______ per hour (1) ______ for each year that he or she ages.

Is age an important determinant of earnings?

- **A.** No, because older workers do not earn significantly more than younger workers.
- **B.** Yes, because wages are not consistent across different age groups.
- **C.** Yes, because younger workers earn more than older workers.
- **D.** No, because the difference in wages is minimal.

Sally is a 25-year-old female college graduate. Betsy is a 43-year-old female college graduate. Predict Sally's and Betsy's earnings.

Sally's earnings prediction is $ ______ per hour. (*Round your response to two decimal places.*)

Betsy's earnings prediction is $ ______ per hour. (*Round your response to two decimal places.*)

The expected difference in earnings between Sally and Betsy is $[   ] per hour. (*Round your response to two decimal places.*)

(1) ○ more
○ less

Answers 0.32

(1) more

B. Yes, because wages are not consistent across different age groups.

15.91

21.67

5.76

ID: Exercise 6.3

---

11. Data were collected from a random sample of 350 home sales from a community in 2003. Let *Price* denote the selling price (in $1,000), *BDR* denote the number of bedrooms, *Bath* denote the number of bathrooms, *Hsize* denote the size of the house (in square feet), *Lsize* denote the lot size (in square feet), *Age* denote the age of the house (in years), and *Poor* denote a binary variable that is equal to 1 if the condition of the house is reported as "poor."

An estimated regression yields

$$\widehat{Price} = 126.4 + 0.514BDR + 24.8Bath + 0.165Hsize + 0.003Lsize$$
$$+ 0.095Age - 51.7Poor,\ \bar{R}^2 = 0.76,\ SER = 44.0.$$

Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?

The expected increase in the value of the house is $[   ]. (*Round your response to the nearest dollar.*)

Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 106 square feet. What is the expected increase in the value of the house?

The expected increase in the value of the house is $[   ]. (*Round your response to the nearest dollar.*)

What is the loss in value if a homeowner lets his house run down so that its condition becomes "poor"?

The loss is $[   ]. (*Round your response to the nearest dollar.*)

Compute the R^2 for the regression.

The R^2 for the regression is [   ]. (*Round your response to three decimal places.*)

Answers 24,800

42,290

51,700

0.764

ID: Exercise 6.5

---

12. Suppose the population multiple regression model is of the form:

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i,$$

where $X_{0i} = 1$, $i = 1, ..., n$.

In the given model, (1) ________ is the constant term and (2) ________ is the constant regressor.

Suppose that a local law enforcement chief wants to decide whether or not to hire more law enforcement officers to patrol area X. For this, the chief wants to know about the current presence of law enforcement officers in area X and the reduction in the incidence of crime in that area. In the last 3 years, the total number of law enforcement officers hired for patrolling area X increased from 1,394 to 1,795. In the same span, the number of crime incidents recorded in area X decreased from 11,522 to 10,155. The relationship the chief wants to estimate is:

$$Reported\ crimes = \beta_0 + \beta_{Officers} \times Officers + \beta_{Unemployed} \times Unemployed + u_{other},$$

where β_0, $\beta_{Officers}$, and $\beta_{Unemployed}$ are the coefficients of the regression line, *Officers* denotes the number of law enforcement officers who were patrolling area X in the last 3 years, *Unemployed* denotes the number of people living in area X who were unemployed for more than 2 months within the last 3 years, and u_{other} is the error term which includes all other factors which could affect the number of reported crimes apart from the presence of law enforcement officers and the number of unemployed people.

From the given information, the partial effect of changing the number of law enforcement officers on the number of reported crimes, $\beta_{Officers}$, holding the number of unemployed individuals fixed, is ☐.

(*Round your answer to two decimal places. Enter a minus sign if your answer is negative.*)

(1)
- ○ X_{0i}
- ○ β_2
- ○ u_i
- ○ β_0
- ○ β_1

(2)
- ○ u_i
- ○ X_{0i}
- ○ X_{1i}
- ○ β_0
- ○ X_{2i}

Answers (1) β_0

(2) X_{0i}

-3.41

ID: Concept Exercise 6.2.1

13. Data on 220 reported crimes is collected from district *X* in 2016. Suppose *CS* denotes the total cost to the state of offering crime protection services to this district (in thousand dollars), *LEOP* denotes the number of law enforcement officers on patrol, *DTP* denotes the damage to public and private property (in thousand dollars), and *Prison* denotes the number of prison inmates. An estimated regression yields:

$$\widehat{CS} = 200.1 + 13.5\ LEOP + 0.59\ DTP + 3.12\ Prison.$$

Suppose that as a result of a major drug cartel bust in the district, the number of inmates in the district's prisons drives up by 50.

The expected change in the cost to the state, keeping other factors constant, would be $_____.

(*Express your answer in dollars.*)

Suppose the researcher realises that she may have missed an important variable while processing the data. She now estimates the following regression equation:

$$\widehat{CS} = 245.25 + 18.5\ LEOP + 0.85\ DTP + 0.7\ CCTV + 1.12\ Prison,$$

where *CCTV* denotes the number of CCTV cameras installed in the district.

With the revised estimates, the expected change in the cost to the state due to the increase in the number of inmates would be $_____.

(*Express your answer in dollars.*)

The initial regression model suffered from (1) __________.

(1) ○ dummy variable trap
○ omitted variable bias
○ perfect multicollinearity

Answers 156,000

56,000

(1) omitted variable bias

ID: Concept Exercise 6.3.1

14. Refer to the table of estimated regressions below, computed using data for 1998 from the CPS, to answer the following question. The data set consists of information on 4000 full-time full-year workers. The highest educational achievement of each worker was either a high school diploma or a bachelor's degree. The worker's age ranged from 25 to 34 years. The data set also contained information on the region of the country where the person lived, marital status, and number of children. A detailed description of the variables used in the data set is available here .

Compute the $\bar{R}^2$ for each of the regressions.

Results of Regressions of Average Hourly Earnings on Gender and Education Binary Variables and Other Characteristics Using 1998 Data from the Curren Population Survey.

Dependent variable: average hourly earnings (AHE).

Regressor	**(1)**	**(2)**	**(3)**
College (X_1)	5.46	5.48	5.44
Female (X_2)	-2.64	-2.62	-2.62
Age (X_3)		0.29	0.29
Northeast (X_4)			0.69
Midwest (X_5)			0.60
South (X_6)			-0.27
Intercept	12.69	4.40	3.75
Summary Statistics			
SER	6.27	6.22	6.21
R^2	0.144	0.147	0.188
$\bar{R}^2$			
n	4829	4829	4829

(*Round your response to three decimal places*)

Answers 0.144

0.146

0.187

ID: Exercise 6.1

15. The data set consists of information on 4100 full-time full-year workers. The highest educational achievement for each worker was either a high school diploma or a bachelor's degree. The worker's ages ranged from 25 to 45 years. The data set also contained information on the region of the country where the person lived, marital status, and number of children. For the purposes of these exercises, let

AHE = average hourly earnings (in 2005 dollars)
College = binary variable (1 if college, 0 if high school)
Female = binary variable (1 if female, 0 if male)
Age = age (in years)
Ntheast = binary variable (1 if Region = Northeast, 0 otherwise)
Midwest = binary variable (1 if Region = Midwest, 0 otherwise)
South = binary variable (1 if Region = South, 0 otherwise)
West = binary variable (1 if Region = West, 0 otherwise)

Results of Regressions of Average Hourly Earnings on Gender and Education Binary Variables and Other Characteristics Using Data from the Current Population Survey

Dependent variable: average hourly earnings (*AHE*).

Regressor	(1)	(2)	(3)
College (X_1)	5.30	5.32	5.28
Female (X_2)	−2.56	−2.54	−2.54
Age (X_3)		0.28	0.28
Northeast (X_4)			0.67
Midwest (X_5)			0.58
South (X_6)			−0.26
Intercept	12.31	4.27	3.64
Summary Statistics			
SER	6.08	6.03	6.02
R^2	0.171	0.184	0.188
$\bar{R}^2$			
n	4100	4100	4100

Compute $\bar{R}^2$ for each of the regressions.

$\bar{R}^2$ for column (1) is ____. (*Round your response to three decimal places.*)

$\bar{R}^2$ for column (2) is ____. (*Round your response to three decimal places.*)

$\bar{R}^2$ for column (3) is ____. (*Round your response to three decimal places.*)

Using the regression results in column (1):

Workers with college degrees earn $____ per hour (1) ________, on average, than workers with only high school degrees. (*Round your response to two decimal places.*)

Men earn $____ per hour (2) ________ than women on average. (*Round your response to two decimal places.*)

(1) ○ more
○ less

(2) ○ more
○ less

Answers 0.171

0.183

0.187

5.30

(1) more

2.56

(2) more

ID: Exercise 6.2

---

16. In the multiple regression model, the adjusted R^2, $\bar{R}^2$:

- A. cannot decrease when an additional explanatory variable is added.
- B. equals the square of the correlation coefficient *r*.
- C. will never be greater than the regression R^2.
- D. cannot be negative.

Answer: C. will never be greater than the regression R^2.

ID: Test A Ex 6.4.2

---

17. The adjusted R^2, or $\bar{R}^2$, is given by:

- A. $1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$.
- B. $1 - \frac{n-1}{n-k-1}\frac{ESS}{TSS}$.
- C. $\frac{ESS}{TSS}$.
- D. $1 - \frac{n-2}{n-k-1}\frac{SSR}{TSS}$.

Answer: A. $1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$.

ID: Test A Ex 6.4.3

---

18. Suppose n and k denote the sample size and number of regressors respectively.

What degrees-of-freedom adjustment is required while calculating the standard error of the regression (*SER*) from the sum of squared residuals (*SSR*)?

- **A.** A multiplication by $n - k$.
- **B.** A division by $n - k - 1$.
- **C.** A division by $n - k + 1$.
- **D.** A multiplication by $nk + 1$.

A group of randomly selected 180 automobile dealers across the automobile industry was surveyed and information about their advertisement expenditure (X_1), the average price of the cars they sell (X_2), and the total number of cars sold in a particular year (Y) (in thousands) was collected. The estimated OLS regression is:

$$\hat{Y}_i = 10.25 + 1.96X_{1i} - 0.36X_{2i} + u_i,$$

where $\hat{Y}_i$, X_{1i}, and X_{2i} denote the predicted value of the number of cars sold, the advertisment expenditure, and the average price of the cars sold by the i^{th} dealer, respectively.

Calculations show that:

$$\sum_{i=1}^{180} \left(Y_i - \hat{Y}_i\right)^2 = 178.14.$$

Based on the above information, the standard error of regression (*SER*) will be ______.

(*Round your answer to two decimal places*).

Answers B. A division by $n - k - 1$.

1.00

ID: Concept Exercise 6.4.1

19. Health insurance companies are generally faced with the problem of how much premium to charge the customers. Generally, the premium charged by the company (P_i) is decided on the basis of the age of the person (A_i), and the duration for which the insurance is taken (D_i). A health insurance company collects random data on 150 customers. The estimated regression function is:

$$\hat{P}_i = 2.5 + 0.80A_i + 0.75D_i.$$

The insurance company makes the following calculations.

$$\sum_{i=1}^{150} \left(\hat{P}_i - \bar{P}\right)^2 = 70.98.$$

$$\sum_{i=1}^{150} \left(P_i - \bar{P}\right)^2 = 125.25.$$

The R^2 for the estimated regression is ______.

(*Round your answer to two decimal places.*)

The health insurance company realizes that it has not included another determinant, the ailments that the customers suffer from, which also affects the premium charged by the company. To include this determinant, the company uses data on customers' expenditure on medicines (E_i). The estimated regression function is:

$$\hat{P}_i = 2.48 + 0.76A_i + 0.68D_i + 0.65E_i.$$

After including the data on expenditure on medicines, the company makes the following calculations.

$$\sum_{i=1}^{150} \left(P_i - \hat{P}_i\right)^2 = 28.75.$$

$$\sum_{i=1}^{150} \left(P_i - \bar{P}\right)^2 = 125.25.$$

The R^2 for this estimated regression is ______.

(*Round your answer to two decimal places.*)

What is the effect on R^2 and *SSR* if the coefficient of the added regressor is exactly 0?

- **A.** If the coefficient of the added regressor is exactly 0, both the R^2 and *SSR* increase.
- **B.** If the coefficient of the added regressor is exactly 0, the R^2 and *SSR* both do not change.
- **C.** If the coefficient of the added regressor is exactly 0, the R^2 increases and the *SSR* decreases.
- **D.** If the coefficient of the added regressor is exactly 0, the R^2 decreases and the *SSR* increases.

Answers 0.57

0.77

B. If the coefficient of the added regressor is exactly 0, the R^2 and *SSR* both do not change.

ID: Concept Exercise 6.4.2

20. An agricultural researcher wants to study last year's crop yield in country A. Generally, crop yield Y_i depends on the amount of rainfall (measured in inches), R_i, and the amount of fertilizers used (in kilograms), F_i. She selects a random sample of 200 farmlands for her study and estimates the following regression function:

$$\hat{Y}_i = 3.5 + 0.75R_i + 0.8F_i.$$

From the sample data she makes the following calculations.

$$\sum_{i=1}^{200} = \left(Y_i - \hat{Y}_i\right)^2 = 35.56.$$

$$\sum_{i=1}^{200} = \left(Y_i - \bar{Y}\right)^2 = 135.25.$$

The $\bar{R}^2$ for the estimated regression is [ ].

(*Round your answer to two decimal places.*)

Which of the following statements is true?

- **A.** $\bar{R}^2$ can be negative when the regressors, taken together, reduce the sum of squared residuals by such a small amount that this reduction fails to offset the factor $\frac{n-1}{n-k-1}$.
- **B.** R^2 can be negative when the regressors, taken together, reduce the sum of squared residuals by such a small amount that this reduction fails to offset the factor $\frac{n-1}{n-k-1}$.
- **C.** $\bar{R}^2$ can be negative when the regressors, taken together, increase the sum of squared residuals by such a huge amount that this increase fails to offset the factor $\frac{n-1}{n-k-1}$.
- **D.** R^2 can be negative when the regressors, taken together, increase the sum of squared residuals by such a huge amount that this increase fails to offset the factor $\frac{n-1}{n-k-1}$.

Answers 0.74

A.
$\bar{R}^2$ can be negative when the regressors, taken together, reduce the sum of squared residuals by such a small amount that this reduction fails to offset the factor $\frac{n-1}{n-k-1}$.

ID: Concept Exercise 6.4.3

21. A professor is interested in understanding the factors on which the grades a student obtains in her final exam depends. She chooses the student's intelligence quotient (X_1), the grades obtained in the mid-term exam (X_2), the number of days the student was present in the class in the whole semester (X_3) as regressors.

She randomly selects 120 students and estimates the following regression function:

$$\hat{Y}_i = 56 + 1.85X_{1i} + 0.45X_{2i} + 2.13X_{3i} + u_i,$$

where $\hat{Y}_i$, X_{1i}, X_{2i}, and X_{3i} denote the predicted value of grades obtained by the i^{th} student, her intellegence quotient, the grades obtained by her in the mid-term exam, and the number of days she was present in the class in the whole semester, respectively.

She calculates the value of the regression R^2 to be 0.54.

The value of the adjusted R^2 ($\bar{R}^2$) will be [ ].

(*Round your answer to two decimal places*).

Which of the following statements is not true about the $\bar{R}^2$? (*Check all that apply.*)

☐ **A.** The value of $\bar{R}^2$ always lies between 0 and 1.

☐ **B.** The value of $\bar{R}^2$ can be negative.

☐ **C.** The value of R^2 is always greater than the $\bar{R}^2$ given that the number of regressors remains the same in the regression equation.

☐ **D.** The value of $\bar{R}^2$ increases when a new regressor is added to the regression equation.

Answers 0.53

A. The value of $\bar{R}^2$ always lies between 0 and 1., D.

The value of $\bar{R}^2$ increases when a new regressor is added to the regression equation.

ID: Concept Exercise 6.4.4

---

22. (Y_i, X_{1i}, X_{2i}) satisfy the following assumptions . You are interested in β_1, the causal effect of X_1 on Y. Suppose that X_1 and X_2 are uncorrelated. You estimate β_1 by regressing X_1 (so that X_2 is not included in the regression). Does this estimator suffer from omitted variable bias?

○ **A.** Yes.

○ **B.** No.

Answer: B. No.

ID: Exercise 6.9

---

23. (Y_i, X_{1i}, X_{2i}) satisfy the following assumptions ; in addition, var$(u_i|X_{1i}, X_{2i}) = 3$ and var$(X_{1i}) = 6$. A random sample of size $n = 309$ is drawn from the population. Use the formula below to answer the following questions,

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left[\frac{1}{1-\rho^2_{X_1, X_2}}\right]\frac{\sigma^2_u}{\sigma^2_{X_1}}$$

Assume that X_1 and X_2 are uncorrelated. Compute the variance of $\hat{\beta}_1$.

The variance of $\hat{\beta}_1$ is,

$$\sigma^2_{\hat{\beta}_1} = \square.$$

(*Round your response to five decimal places*)

Assume that corr$(X_1, X_2) = 0.37$. Compute the variance of $\hat{\beta}_1$.

The variance of $\hat{\beta}_1$ is,

$$\sigma^2_{\hat{\beta}_1} = \square.$$

(*Round your response to five decimal places*)

Which of the following statements is correct?

- **A.** The larger the correlation between X_1 and X_2, the larger the variance of $\hat{\beta}_1$. Thus, it is best to leave X_2 out of the regression even if it is a determinant of Y.
- **B.** The larger the correlation between X_1 and X_2, the smaller the variance of $\hat{\beta}_1$. Nevertheless, it is best include X_2 in the regression if it is a determinant of Y.
- **C.** The larger the correlation between X_1 and X_2, the smaller the variance of $\hat{\beta}_1$. Thus, it is best to leave X_2 out of the regression even if it is a determinant of Y.
- **D.** The larger the correlation between X_1 and X_2, the larger the variance of $\hat{\beta}_1$. Nevertheless, it is best include X_2 in the regression it is a determinant of Y.

Answers 0.00162

0.00187

D.

The larger the correlation between X_1 and X_2, the larger the variance of $\hat{\beta}_1$. Nevertheless, it is best include X_2 in the regression if it is a determinant of Y.

ID: Exercise 6.10

24. In which of the following scenarios does perfect multicollinearity occur?

- **A.** Perfect multicollinearity occurs when the value of kurtosis for the dependent and explanatory variables is infinite.
- **B.** Perfect multicollinearity occurs when one of the regressors is a perfect linear function of the other regressors.
- **C.** Perfect multicollinearity occurs when the regressors are independently and identically distributed.
- **D.** Perfect multicollinearity occurs when one of the regressors is an exponential function of the other regressors.

Why is it impossible to compute OLS estimators in the presence of perfect multicollinearity?

- **A.** It is impossible to compute OLS estimators in the presence of perfect multicollinearity because adding a new explanatory variable does not increase fit of the model.
- **B.** It is impossible to compute OLS estimators in the presence of perfect multicollinearity because it produces multiplication by 0.
- **C.** It is impossible to compute OLS estimators in the presence of perfect multicollinearity because the conditional distribution of the error term has a nonzero mean.
- **D.** It is impossible to compute OLS estimators in the presence of perfect multicollinearity because it produces division by 0.

Perfect multicollinearity can be rectified by modifying the (1) ____________.

(1)
- dependent variable
- independent variables
- error term

Answers B. Perfect multicollinearity occurs when one of the regressors is a perfect linear function of the other regressors.

D. It is impossible to compute OLS estimators in the presence of perfect multicollinearity because it produces division by 0.

(1) independent variables

ID: Concept Exercise 6.5.1

---

25. Under the least squares assumptions for the multiple regression problem (zero conditional mean for the error term, all X_i and Y_i being i.i.d., all X_i and μ_i having finite fourth moments, no perfect multicollinearity), the OLS estimators for the slopes and intercept:

- **A.** have an exact normal distribution for $n > 25$.
- **B.** are BLUE.
- **C.** are unbiased and consistent.
- **D.** have a normal distribution in small samples as long as the errors are homoskedastic.

Answer: C. are unbiased and consistent.

ID: Test A Ex 6.6.4

---

26. The following OLS assumption is most likely violated by omitted variables bias:

- **A.** $E(\mu_i | X_i) = 0$.
- **B.** there are no outliers for X_i, μ_i.
- **C.** (X_i, Y_i), $i = 1,\ldots,n$ are i.i.d. draws from their joint distribution.
- **D.** there is heteroskedasticity.

Answer: A. $E(\mu_i | X_i) = 0$.

ID: Test B Ex 6.6.3

---

27. Consider the following multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

Which of the following explains why two perfectly collinear regressors cannot be included in a linear multiple regressions? (*Check all that apply*)

- **A.** For the case of two regressors and homoskedasticity, it can be shown mathematically that the variance of the estimated coefficient $\widehat{\beta_1}$ goes to infinity as the correlation between X_1 and X_2 goes to one.
- **B.** Perfectly collinear regressors cannot be included in a linear multiple regression because it shrinks the estimated coefficients $\widehat{\beta_1}$ and $\widehat{\beta_2}$ towards zero.
- **C.** Intuitively, if one regressor is a linear function of another, OLS cannot identify the partial effect of one while holding the other constant.
- **D.** None of the above are correct.

Suppose you are interested in estimating the effect of the student–teacher ratio (*STR*) on test performance using the model

$$TestScore = \beta_0 + \beta_1 STR + u$$

Which of the following regressors, if added to the model, would be perfectly collinear with *STR*? (*Check all that apply*)

- **A.** The number of teachers expressed as a percentage of the number of students.
- **B.** The teacher–security officer ratio if the student–security officer ratio is 1:10.
- **C.** An indicator variable that equals one if a student is a native English speaker.
- **D.** The student–faculty parking ratio, if every teacher has a parking spot.

Answers A.

For the case of two regressors and homoskedasticity, it can be shown mathematically that the variance of the estimated coefficient $\widehat{\beta_1}$ goes to infinity as the correlation between X_1 and X_2 goes to one.

, C.

Intuitively, if one regressor is a linear function of another, OLS cannot identify the partial effect of one while holding the other constant.

A. The number of teachers expressed as a percentage of the number of students., B.
The teacher–security officer ratio if the student–security officer ratio is 1:10., D.
The student–faculty parking ratio, if every teacher has a parking spot.

ID: Review Concept 6.3

---

28. Consider the following multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

Which of the following explains why it is difficult to estimate precisely the partial effect of X_1, holding X_2 constant, if X_1 and X_2 are highly correlated?

☐ **A.** Perfectly collinear regressors cannot be included in a linear multiple regression because it shrinks the estimated coefficients $\widehat{\beta_1}$ and $\widehat{\beta_2}$ towards zero.

☐ **B.** For the case of two regressors and homoskedasticity, it can be shown mathematically that the variance of the estimated coefficient $\widehat{\beta_1}$ goes to infinity as the correlation between X_1 and X_2 goes to one.

☐ **C.** Intuitively, estimating the partial effect of X_1 holding X_2 constant becomes difficult because after controlling for X_2, there is little variation left to precisely estimate the partial effect of X_1.

☐ **D.** None of the above are correct.

Answer: B.

For the case of two regressors and homoskedasticity, it can be shown mathematically that the variance of the estimated coefficient $\widehat{\beta_1}$ goes to infinity as the correlation between X_1 and X_2 goes to one.

, C.

Intuitively, estimating the partial effect of X_1 holding X_2 constant becomes difficult because after controlling for X_2, there is little variation left to precisely estimate the partial effect of X_1.

ID: Review Concept 6.4

29. The data set consists of information on 3600 full-time full-year workers. The highest educational achievement for each worker was either a high school diploma or a bachelor's degree. The worker's ages ranged from 25 to 45 years. The data set also contained information on the region of the country where the person lived, marital status, and number of children. For the purposes of these exercises, let

AHE = average hourly earnings (in 2005 dollars)
College = binary variable (1 if college, 0 if high school)
Female = binary variable (1 if female, 0 if male)
Age = age (in years)
Ntheast = binary variable (1 if Region = Northeast, 0 otherwise)
Midwest = binary variable (1 if Region = Midwest, 0 otherwise)
South = binary variable (1 if Region = South, 0 otherwise)
West = binary variable (1 if Region = West, 0 otherwise)

Results of Regressions of Average Hourly Earnings on Gender and Education Binary Variables and Other Characteristics Using Data from the Current Population Survey

Dependent variable: average hourly earnings (*AHE*).

Regressor	(1)	(2)	(3)
College (X_1)	5.30	5.32	5.28
Female (X_2)	−2.56	−2.54	−2.54
Age (X_3)		0.28	0.28
Northeast (X_4)			0.67
Midwest (X_5)			0.58
South (X_6)			−0.26
Intercept	12.31	4.27	3.64
Summary Statistics			
SER	6.08	6.03	6.02
R^2	0.171	0.184	0.188
$\bar{R}^2$	0.171	0.183	0.187
n	3600	3600	3600

Using the regression results in column (3):

Workers in the Northeast earn $____ more per hour than workers in the West, on average, controlling for other variables in the regression. (*Round your response to two decimal places.*)

Workers in the South earn $____ less per hour than workers in the West, on average, controlling for other variables in the regression. (*Round your response to two decimal places.*)

Do there appear to be important regional differences?

- **A.** No, because the difference in wages is minimal.
- **B.** No, because workers in the Midwest earn more per hour than workers in the Northeast, while workers in the South earn less per hour than workers in the West.
- **C.** Yes, because workers in the Northeast earn significantly more than workers in the South.
- **D.** Yes, because wages are not consistent across the region.

Why is the regressor *West* omitted from the regression? What would happen if it was included?

- **A.** The regressor *West* is omitted to avoid the OLS estimator bias. If *West* is included, then the

OLS estimator is biased.

- **B.** The regressor *West* is omitted to avoid perfect multicollinearity. If *West* is included, then the OLS estimator cannot be computed in this situation.
- **C.** The regressor *West* is omitted to make the OLS estimator linear. If *West* is included, then the intercept can no longer be written as a perfect linear function of the four regional regressors.
- **D.** The regressor *West* is omitted to minimize the sum of squared residuals. If *West* is included, then the sum of squared residuals is no longer minimized.

Juanita is a 30-year-old female college graduate from the South. Jennifer is a 30-year-old female college graduate from the Midwest. Calcu the expected difference in earnings between Juanita and Jennifer.

The expected difference in earnings between Juanita and Jennifer is $ ______ per hour. (*Round your response to two decimal places.*)

Answers 0.67

0.26

D. Yes, because wages are not consistent across the region.

B.

The regressor *West* is omitted to avoid perfect multicollinearity. If *West* is included, then the OLS estimator cannot be computed this situation.

0.84

ID: Exercise 6.4

30. This exercise will use the Excel data set, Growth[1], described[2] in Empirical Exercise 4.4. A description of the variables in provided in the Word file, Growth_Description.

Obtain the mean and standard deviation of the series, *Growth*, *TradeShare*, *YearsSchool*, *Oil*, *Rev_Coups*, *Assassinations* and *RGDP60*. The standard deviation of *RGDP60* is

- **A.** 3131
- **B.** 2513
- **C.** 3103
- **D.** 2523

Regress Growth on *TradeShare*, *YearsSchool*, *Rev_Coups*, *Assassinations* and *RGDP60*. Interpret the coefficient on *YearsSchool* holding the other variables constant.

- **A.** Every one additional year of completed adult schooling decreases the average annual growth rate by 0.58 percentage points
- **B.** Every one additional year of completed adult schooling increases the average annual growth rate by 0.58 percentage points
- **C.** Every one additional year of completed adult schooling increases the average annual growth rate by 58 percent
- **D.** Every one additional year of completed adult schooling increases the average annual growth rate by 5.8 percent

Is the coefficient on *YearsSchool* large or small?

- **A.** Cannot be determined
- **B.** It is rather small given that today most adults have at least 12 years of schooling and one additional year will not affect a country's growth rate
- **C.** It is NOT large because one year of schooling, a 25 percent increase evaluated at the mean of schooling, increases annual growth by .58 percentage points over a mean growth rate of 1.9 percentage points which is inconsequential
- **D.** It is large because one year of schooling, a 25 percent increase evaluated at the mean of schooling, increases annual growth by .58 percentage points over a mean growth rate of 1.9 percentage points, and thus by more than 25 percent

Using the regression results in part (2), what is the expected growth rate of a country evaluated at the mean of all the regressors?

- **A.** 1.87%
- **B.** 1.94 %
- **C.** 2.87%
- **D.** 1.17%

Repeat (4) but now assume that the country's value for *YearsSchool* is one standard deviation above its mean. The expected growth rate would be

- **A.** 1.94 %
- **B.** 3.15 %
- **C.** 4.24 %
- **D.** 3.40 %

Notice that Malta's trade share is very large, almost double its GDP. This seems extreme. To assess the sensitivity of your results to the inclusion of Malta, redo the regression in part (2) but drop the observation on Malta (now n=64 instead of n=65). How did the coefficient on *TradeShares* change?

- **A.** All the above
- **B.** It fell about 14 percent

C. Without Malta a one percentage point increase in the *Tradeshares* will increase *Growth* by 0.0134 percentage points

D. It fell from 1.56 to 1.34

Re-run the regression in part (2) but add the variable *Oil* to the model. What is coefficient on *Oil?*

A. None of the above

B. 0.0 because oil is perfectly collinear with the constant term

C. 0.0 because oil is not an important predictor of growth

D. 0.0 because oil is perfectly collinear with *Tradeshares*

The problem with including a variable like *Oil* is that it leads to

A. the dummy variable trap if an intercept is included

B. A change in the other coefficients

C. None of the above

D. An insignificant coefficient because trade in oil is unrelated to GDP

Compare the unadjusted R^2 from the regressions in part (2) and part (7). Why are they the same?

A. Because *oil* adds some but not enough explanatory power to the model

B. Because the coefficient on *oil* is too small to matter

C. Because the effect of *oil* cannot be estimated since it doesn't vary in this sample of countries

D. It is just a coincidence

Compare the *adjusted* R^2 from the regressions in part (2) and part (7). Why are they NOT the same?

A. Because, k, the number of regressors in the two models is different

B. Because, n, the number of observations in the two models is different

C. Because the adjusted R^2 falls if you add regressors that have no explanatory power

D. Both a and c

1: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/Growth.xls

2: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/Growth_Description.pdf

Answers B. 2513

B. Every one additional year of completed adult schooling increases the average annual growth rate by 0.58 percentage points

D.
It is large because one year of schooling, a 25 percent increase evaluated at the mean of schooling, increases annual growth by .58 percentage points over a mean growth rate of 1.9 percentage points, and thus by more than 25 percent

B. 1.94 %

D. 3.40 %

A. All the above

B. 0.0 because oil is perfectly collinear with the constant term

A. the dummy variable trap if an intercept is included

C. Because the effect of *oil* cannot be estimated since it doesn't vary in this sample of countries

D. Both a and c

ID: Empirical Exercise 6.2 (static)

31. Using the Excel data set, CollegeDistance[3], described[4] in Empirical Exercise 4.3, run a regression of years of completed schooling (*ed*) on distance (in 10s of miles) from a 4-year college (*dist*).

The coefficient on distance (*dist*) shows the

- **A.** Years of completed schooling increase by 0.073 years for every 1-mile increase in distance from the nearest 4-year college
- **B.** Years of completed schooling increase by 0.73 years for every 100-mile increase in distance from the nearest 4-year college
- **C.** Years of completed schooling decrease by 0.073 years for every 10-mile increase in distance from the nearest 4-year college
- **D.** Years of completed schooling increase by 0.073 years for every 10-mile increase in distance from the nearest 4-year college

Regress completed schooling (*ed*) on the variables *dist*, *female*, *black*, *Hispanic*, *bytest*, *dadcoll incomehi*, *ownhome*, *cue80*, and stwmfg80. The coefficient on distance (*dist*) now indicates that adjusted for other factors,

- **A.** Years of completed schooling decrease by 0.032 years for every 1-mile increase in distance from the nearest 4-year college
- **B.** Years of completed schooling decrease by 0.032 years for every 10 miles closer one lives from the nearest 4-year college
- **C.** Years of completed schooling decrease by 0.32 years for every 10-mile increase in distance from the nearest 4-year college
- **D.** Years of completed schooling increase by 0.032 years for every 10 miles closer one lives from the nearest 4-year college

Based on a comparison of the coefficients on *dist* in the regressions in part (1) and part (2) we would conclude that

- **A.** There was no omitted variable problem because distance to a 4-year college is independent of the other factors
- **B.** There is an omitted variable problem because the R^2 in the multiple regression in part (2) is much greater than the R^2 in the simple regression of part (1)
- **C.** There is no omitted variable problem with the regression in part (1)
- **D.** There is an omitted variable problem because the coefficient on distance was reduced by 57% suggesting that other factors correlated with distance but also correlated with completed schooling were not included in the simple regression

Based on the multiple regression results in part (2), we could say that

- **A.** 27.9% of the variation in schooling is explained by the variables
- **B.** None of the above
- **C.** The additional variables added to the regression in part (2) do not add much explanatory power relative to the model in part (1)
- **D.** 2.79% of the variation in schooling is explained by the variables

The coefficient on female in part (2) indicates that

- **A.** Females obtain 0.145 fewer years of schooling than do males adjusted for the other factors
- **B.** Females obtain 0.145 more years of schooling than do males adjusted for the other factors
- **C.** Females obtain 1.45 more years of schooling than do males adjusted for the other factors
- **D.** Females obtain 1.45 fewer years of schooling than do males adjusted for the other factors

How many years of schooling would a black female be expected to have if she has a base year test score of 50; a father that went to college; is from a family with income greater than $25,000 that owns its home; she is from a county where the unemployment rate is 6.0, and the state hourly wage in manufacturing is $8.00 and she lives 100 miles from the nearest 4-year college?

A. 14.68
B. 12.82
C. None of the above
D. 8.83

How many years of schooling would a black female be expected to have if she had the same characteristic as in part (7) but her family had less than $25,000 in income and they did not own their own home?

A. 15.08
B. 15.6
C. 14.14
D. 12.08

How many years of schooling would a person be expected to have if all you knew was that they lived 100 miles from the nearest 4-year college (Hint you must use the results from part (1)).

A. 12.82
B. 13.96
C. 15.62
D. 13.22

The coefficient on *DadColl* from the regression in part (2) indicates that

A. This person would be expected to have 6.96 more years of schooling than the same person whose father did not have a college degree
B. This person would be expected to have 0.696 fewer years of schooling than the same person whose father did not have a college degree
C. This person would be expected to have 0.696 more years of schooling than the same person whose father did not have a college degree
D. This person would be expected to have 0.069 more years of schooling than the same person whose father did not have a college degree

3: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/CollegeDistance.xls

4: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/CollegeDistance_DataDescripti

Answers C. Years of completed schooling decrease by 0.073 years for every 10-mile increase in distance from the nearest 4-year college

D. Years of completed schooling increase by 0.032 years for every 10 miles closer one lives from the nearest 4-year college

D.
There is an omitted variable problem because the coefficient on distance was reduced by 57% suggesting that other factors correlated with distance but also correlated with completed schooling were not included in the simple regression

A. 27.9% of the variation in schooling is explained by the variables

B. Females obtain 0.145 more years of schooling than do males adjusted for the other factors

A. 14.68

C. 14.14

D. 13.22

C.
This person would be expected to have 0.696 more years of schooling than the same person whose father did not have a colleg degree

ID: General Empirical 6.1 (static)

---

32. The dummy variable trap is an example of:

- **A.** imperfect multicollinearity.
- **B.** perfect multicollinearity.
- **C.** something that is of theoretical interest only.
- **D.** something that does not happen to university or college students.

Answer: B. perfect multicollinearity.

ID: Test A Ex 6.7.5

---

33. Imperfect multicollinearity:

- **A.** means that the least squares estimator of the slope is biased.
- **B.** is not relevant to the field of economics and business administration.
- **C.** only occurs in the study of finance.
- **D.** means that two or more of the regressors are highly correlated.

Answer: D. means that two or more of the regressors are highly correlated.

ID: Test B Ex 6.7.4

---

34. Imperfect multicollinearity:

- **A.** implies that it will be difficult to estimate precisely one or more of the partial effects using the data at hand.
- **B.** violates one of the four Least Squares assumptions in the multiple regression model.
- **C.** means that you cannot estimate the effect of at least one of the *X*s on *Y*.
- **D.** suggests that a standard spreadsheet program does not have enough power to estimate the multiple regression model.

Answer: A. implies that it will be difficult to estimate precisely one or more of the partial effects using the data at hand.

ID: Test B Ex 6.7.5

---

35. A real estate agency wants to understand the factors affecting the price of houses which are up for re-selling. The study includes 1,000 houses and estimates the following regression using OLS:

$$Price = \beta_{age}X_{age} + \beta_{numsold}X_{numsold} + \beta_{area}X_{area} + \beta_{garage}X_{garage} + \beta_{citycenter}X_{citycenter} + \beta_{outskirts}X_{outskirts} + \beta_{suburban}X_{suburban} + u_{other}$$

where X_{age} denotes the age of the house, $X_{numsold}$ denotes the number of times the house was sold, X_{area} denotes floor area, X_{garage} denotes the size of the garage, $X_{citycenter}$, $X_{outskirts}$, and $X_{suburban}$ are dummy variables indicating the respective locality of the house: city center, outskirts, or suburban, and u_{other} denotes the error term. Assume that the least squares assumptions are satisfied unless the above information suggests otherwise.

Which of the following statements are true about the explanatory variables used in this study? (*Check all that apply.*)

- **A.** The regressors in this model are perfectly multicollinear due to the absence of the intercept term.
- **B.** The regressors in this model are perfectly multicollinear due to the variables indicating the locality of the house.
- **C.** The estimated coefficients will be jointly normally distributed.
- **D.** The regression does not fall into the dummy variable trap due to the absence of the intercept term.

Suppose that the model now includes an intercept term, β_0.

The model should now include (1) ____________ dummy variables indicating the locality of a house.

Suppose the agency finds a high positive correlation between X_{garage} and X_{area}.

The data set would provide little information about what happens to the prices of houses with a large floor area and a

(2) ____________ garage, or vice versa, keeping other variables constant.

(1)
- any two
- any one
- all three

(2)
- small
- large

Answers C. The estimated coefficients will be jointly normally distributed., D.
The regression does not fall into the dummy variable trap due to the absence of the intercept term.

(1) any two

(2) small

ID: Concept Exercise 6.7.1

---

36. Suppose the population regression function is given by:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i,\ i = 1,..., n,$$

where u_i is the error term and $\beta_1,..., \beta_k$, are the coefficients on the X's.

Which of the following statements is not true about the omitted variable bias?

- **A.** When there is omitted variable bias, the OLS estimator is inconsistent.
- **B.** When there is omitted variable bias, the OLS estimator is biased.
- **C.** Omitted variable bias violates the assumption that X_{1i}, ..., X_{ki} are i.i.d.
- **D.** Omitted variable bias persists even in large samples.

An independent researcher estimates the relationship between the rent a family pays (Y_i) and the size of the house (X_i) (measured in sq. ft.) by surveying 200 prospective locations. The estimated regression function is as follows:

$$\hat{Y}_i = 0.35 + 1.2X_i,$$

where $\hat{Y}_i$ and X_i denote the predicted value of the rent the i^{th} family pays and the size of that house, respectively.

Given that she has included the size of the house as the only regressor for estimating the above equation there might be other variables which can be a determinant of the total rent that a family pays.

In which of the following cases might there arise an omitted variable bias due to the exclusion of explanatory variables?

- **A.** The number of years since the house was constructed.
- **B.** The number of rooms in the house.
- **C.** Whether the house is furnished or not furnished.
- **D.** Distance of the house from the airport.

Let the population regression function with k variables of interest be denoted by X, and let there be r control variables denoted by W be written as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \beta_{k+1} W_{1i} + ... + \beta_{k+r} W_{ri} + u_i,$$

where u_i is the error term and the coefficients on the X's, $\beta_1,..., \beta_k$, are causal effects of interest.

Which of the following is the mathematical statement for the conditional mean independence assumption?

- **A.** $E\left(Y_i|X_{1i},...,X_{ki},W_{1i},...,W_{ri}\right) = E\left(Y_i|X_{1i},...,X_{ri}\right)$.
- **B.** $E\left(u_i|X_{1i},...,X_{ki},W_{1i},...,W_{ri}\right) = E\left(u_i|X_{1i},...,X_{ri}\right)$.
- **C.** $E\left(u_i|X_{1i},...,X_{ki},W_{1i},...,W_{ri}\right) = E\left(u_i|W_{1i},...,W_{ri}\right)$.
- **D.** $E\left(Y_i|X_{1i},...,X_{ki},W_{1i},...,W_{ri}\right) = E\left(Y_i|W_{1i},...,W_{ri}\right)$.

Answers C. Omitted variable bias violates the assumption that X_{1i}, ..., X_{ki} are i.i.d.

B. The number of rooms in the house.

C. $E\left(u_i|X_{1i},...,X_{ki},W_{1i},...,W_{ri}\right) = E\left(u_i|W_{1i},...,W_{ri}\right)$.

ID: Concept Exercise 6.8.1