Student: ____________________	**Instructor:** Richeng Piao	**Assignment:** Practice Problem Set 7
Date: ____________________	**Course:** ECON 2560 - Applied Econometrics	

1. The data set consists of information on 3800 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)	5.35 (0.21)	5.37 (0.21)	5.33 (0.21)
Female (X_2)	−2.59 (0.20)	−2.57 (0.20)	−2.57 (0.20)
Age (X_3)		0.28 (0.04)	0.28 (0.04)
Northeast (X_4)			0.68 (0.29)
Midwest (X_5)			0.59 (0.27)
South (X_6)			−0.26 (0.25)
Intercept	12.44 (0.14)	4.31 (1.03)	3.68 (1.04)
Summary Statistics			
F-statistic for regional effects = 0			6.01
SER	6.14	6.10	6.09
R^2	0.172	0.186	0.190
n	3800	3800	3800

Using the regression results in column (1):

The *t*-statistic for the college–high school earnings difference estimated from this regression is ______. (*Round your response to two decimal places.*)

Is the college–high school earnings difference estimated from this regression statistically significant at the 5% level?

Since the absolute value of the *t*-statistic is (1) ____________ than the critical value for 95% confidence, the college–high school earnings difference estimated from this regression (2) ____________ statistically significant at the 5% level.

Construct a confidence interval of 95% for the college–high school earnings difference.

The 95% confidence interval for the college–high school earnings difference is (____ , ____). (*Round your responses t two decimal places.*)

The *t*-statistic for the male–female earnings difference estimated from this regression is ____ . (*Round your response to two decin places.*)

Is the male–female earnings difference estimated from this regression statistically significant at the 5% level?

Since the *t*-statistic is (3) ____________ than the critical value for 95% confidence, the male–female earnings difference estimated from th regression (4) ____________ statistically significant at the 5% level.

Construct a confidence interval of 95% for the male–female earnings difference.

The 95% confidence interval for the male–female earnings difference is (____ , ____). (*Round your responses to two decimal places.*)

(1) ○ greater ○ less
(2) ○ is ○ is not
(3) ○ greater ○ less
(4) ○ is ○ is not

Answers 25.48

(1) greater

(2) is

4.94

5.76

−12.95

(3) greater

(4) is

−2.98

−2.20

ID: Exercise 7.2

2. The data set consists of information on 3300 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)	5.90 (0.23)	5.92 (0.23)	5.88 (0.23)
Female (X_2)	−2.85 (0.22)	−2.83 (0.22)	−2.83 (0.22)
Age (X_3)		0.31 (0.04)	0.31 (0.05)
Northeast (X_4)			0.75 (0.32)
Midwest (X_5)			0.65 (0.30)
South (X_6)			−0.29 (0.28)
Intercept	13.71 (0.15)	4.75 (1.13)	4.05 (1.14)
Summary Statistics			
F-statistic for regional effects = 0			6.25
SER	6.77	6.72	6.71
R^2	0.190	0.205	0.210
n	3300	3300	3300

Using the regression results in column (2):

The *t*-statistic for the coefficient on *Age* is _____. (*Round your response to two decimal places.*)

The *p*-value for the preceding *t*-statistic is _____. (*Round your response to four decimal places.*)

Does this imply that age is an important determinant of earnings?

- **A.** No, age is not an important determinant of earnings because the high *p*-value implies that the coefficient on age is not statistically s
- **B.** No, age is not an important determinant of earnings because the low *p*-value implies that the coefficient on age is not statistically si
- **C.** Yes, age is an important determinant of earnings because the low *p*-value implies that the coefficient on age is statistically significar
- **D.** Yes, age is an important determinant of earnings because the high *p*-value implies that the coefficient on age is statistically significa

Using the regression results in column (3):

Sally is a 32-year-old female college graduate. Betsy is a 40-year-old female college graduate. Construct a confidence interval of 99% for th expected difference between their earnings.

The 99% confidence interval for the expected difference between their earnings is (,). (*Round your respons to two decimal places.*)

Answers 7.75

0.0000

C.
Yes, age is an important determinant of earnings because the low *p*-value implies that the coefficient on age is statistically significant at the 1% level.

1.45

3.51

ID: Exercise 7.3

---

3. If you wanted to test, using a 5% significance level, whether or not a specific slope coefficient is equal to one, then you should:

- **A.** subtract 1 from the estimated coefficient, divide the difference by the standard error, and check if the resulting ratio is larger than 1.9
- **B.** see if the slope coefficient is between 0.95 and 1.05.
- **C.** add and subtract 1.96 from the slope and check if that interval includes 1.
- **D.** check if the adjusted R^2 is close to 1.

Answer: A.
subtract 1 from the estimated coefficient, divide the difference by the standard error, and check if the resulting ratio is larger than 1.96.

ID: Test A Ex 7.1.1

---

4. Consider the following regression output where the dependent variable is test scores and the two explanatory variables are the student-teacher ratio and the percent of English learners: $\widehat{TestScore} = 698.9 - 1.10 \times STR - 0.650 \times PctEL$. You are told that the *t*-statistic on the student-teacher ratio coefficient is 2.56.

The standard error therefore is approximately:

- **A.** 1.96.
- **B.** 0.43.
- **C.** 0.650.
- **D.** 0.25.

Answer: B. 0.43.

ID: Test A Ex 7.1.2

---

5. In the multiple regression model, the *t*-statistic for testing that the slope is significantly different from zero is calculated:

- **A.** from the square root of the *F*-statistic.
- **B.** by dividing the estimate by its standard error.
- **C.** using the adjusted R^2 and the confidence interval.
- **D.** by multiplying the *p*-value by 1.96.

Answer: B. by dividing the estimate by its standard error.

ID: Test B Ex 7.1.1

---

6. A medical student at a community college in city *Q* wants to study the factors affecting the systolic blood pressure of a person (*Y*). Generally, the systolic blood pressure depends on the BMI of a person (*B*) and the age of the person *A*. She wants to test whether or not the BMI has a significant effect on the systolic blood pressure, keeping the age of the person constant. For her study, she collects a random sample of 125 patients from the city and estimates the following regression function:

$$\hat{Y} = 15.50 + \underset{(0.50)}{0.90B} + \underset{(0.40)}{1.10A}.$$

The test statistic of the study the student wants to conduct ($H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$), keeping other variables constant is ______.

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

At the 5% significance level, the student will (1) ______ the null hypothesis.

Keeping BMI constant, she now wants to test whether the age of a person (*A*) has no significant effect or a positive effect on the person's systolic blood pressure.

So, the test statistic associated with the one-sided test the student wishes to conduct will be ______.

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

At the 5% significance level, the student will (2) ______ the null hypothesis.

(1) reject	(2) fail to reject
fail to reject	reject

Answers 1.80

(1) fail to reject

2.75

(2) reject

ID: Concept Exercise 7.1.1

---

7. A researcher wants to study the performance of high school students of district W in the mathematics final exam. To study the performance, he uses the marks scored by the students in mathematics M as a function of the average number of hours spent by the student on practice H and number of days the student attended the mathematics class in school D. For his study, he selects a random sample of 200 high school students and estimates the following regression function:

$$\widehat{M} = 10.25 + 1.20H + 2.75D.$$

The researcher wants to test whether or not changing the average number of hours spent by the student on practice has a statistically significant impact on the marks scored by the students in mathematics.

Keeping the other variables constant, the null and the alternative hypotheses of the test conducted by the researcher are $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.

The test statistic for the test is 1.25.

Therefore, the *p*-value of the test is ______.

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

If the study uses two-sided test with the 5% significance level, then the *p*-value suggests that we (1) ______ the null hypothesis.

The researcher now wants to test whether or not increasing the number of days the student attended the mathematics class in school significantly improves the marks scored by the students in mathematics.

Keeping the other variables constant, the researcher wants to calculate the *p*-value for the test $H_0: \beta_2 = 0$ vs. $H_1: \beta_2 > 0$.

The test statistic for this test is 2.55.

Therefore, the *p*-value of the test is ______.

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

If the study uses one-sided test with the 5% significance level, then the *p*-value suggests that we (2) ______ the null hypothesis.

(1) ○ reject
○ fail to reject

(2) ○ reject
○ fail to reject

Answers 0.21

(1) fail to reject

0.01

(2) reject

ID: Concept Exercise 7.1.2

8. An independent researcher wants to investigate if the factors which determine the house rent (Y, measured in dollars), such as the distance of the house from the airport (X_1), the time since the house was built (X_2), are significant or not. He collects data from 120 prospective locations and estimates the following regression equation:

$$\hat{Y}_i = 2.5 \underset{(2.45)}{- 0.88X_1} \underset{(1.84)}{+ 2.34X_2}.$$

The 95% confidence interval for the slope coefficient β_1, keeping the other variables constant will be (___ , ___).

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

Based on the calculated confidence intervals, we can say that at the 5% significance level, we will (1) ___ the hypothesis $\beta_1 = 0$.

The 99% confidence interval for the slope coefficient β_2, keeping the other variables constant will be (___ , ___).

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

Based on the calculated confidence intervals, we can say that at the 1% significance level, we will (2) ___ the hypothesis $\beta_2 = 0$.

(1) ○ reject
○ fail to reject

(2) ○ fail to reject
○ reject

Answers −5.68

3.92

(1) fail to reject

−2.41

7.09

(2) fail to reject

ID: Concept Exercise 7.1.3

9. An independent researcher is interested in studying the factors that affects the educational attainment of an individual (Y). For her study she chooses educational attainment of the individual's mother (X_1) and the annual income of the individual's family (X_2) as regressors. The researcher randomly selects a sample of 250 individuals and estimates the following regression function:

$$\hat{Y}_i = \underset{}{10.00} + \underset{(0.56)}{1.74}X_1 + \underset{(1.23)}{1.12}X_2.$$

She wants to test whether or not a change in (X_1) significantly affects (Y) keeping the annual income of the individual's family constant.

The *t*-statistic for testing the hypothesis $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ will be ______.

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

If the researcher uses a 5% significance level, we (1) ______ the null hypothesis.

She now includes the educational attainment of the individual's father (X_3) as the third regressor in the regression equation.

The newly estimated regression equation is:

$$\hat{Y}_i = 8.51 + \underset{(1.75)}{1.76}X_1 + \underset{(2.14)}{2.01}X_2 + \underset{(0.71)}{1.12}X_3.$$

With the revised estimates she again tests the hypothesis $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ keeping the annual income of the individual's family and the educational attainment of the individual's father constant.

The *t*-statistic for the test the researcher wants to conduct will be ______.

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

Based on the value of the test statistic at the 5% significance level, we (2) ______ the null hypothesis.

(1) ○ fail to reject
○ reject

(2) ○ fail to reject
○ reject

Answers 3.11

(1) reject

1.01

(2) fail to reject

ID: Concept Exercise 7.1.4

10. Data were collected from a random sample of 220 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:

$$\widehat{Price} = \underset{(24.1)}{120.4} + \underset{(2.38)}{0.490}BDR + \underset{(9.03)}{23.6}Bath + \underset{(0.011)}{0.158}Hsize + \underset{(0.00048)}{0.004}Lsize$$

$$+ \underset{(0.314)}{0.091}Age - \underset{(10.6)}{49.3}Poor, \; \bar{R}^2 = 0.73, \; SER = 41.9$$

The *t*-statistic for the coefficient on *BDR* is [ ]. (*Round your response to three decimal places.*)

Is the coefficient on *BDR* statistically significantly different from zero?

- **A.** Since the *t*-statistic > 1.96, the coefficient on *BDR* is statistically significantly different from zero.
- **B.** Since the *t*-statistic > 0.05, the coefficient on *BDR* is not statistically significantly different from zero.
- **C.** Since the *t*-statistic < 1.96, the coefficient on *BDR* is not statistically significantly different from zero.
- **D.** Since the *t*-statistic < 0.05, the coefficient on *BDR* is statistically significantly different from zero.

Typically five-bedroom houses sell for much more than two-bedroom houses. Is this consistent with the regression?

- **A.** Yes, the coefficients on *BDR* and *Hsize* accurately take into account that each additional bedroom changes not only the total num
- **B.** No, the coefficient on *BDR* measures the partial effect of the number of bedrooms, holding house size constant, and thus significa

A homeowner purchases 2020 square feet from an adjacent lot. Construct a confidence interval of 95% for the change in the value of her house.

The 95% confidence interval for the effect of lot size on price is ([ ], [ ]) (*in thousands of dollars*). (*Round your responses to two decimal places.*)

Lot size is measured in square feet. Do you think that another scale might be more appropriate?

- **A.** No, choosing another scale would not affect the regression results because the estimate coefficient would remain unaffected.
- **B.** No, if the lot size were measured in thousands of square feet, the estimate coefficient would be 0.000004 instead of 0.004, thus m
- **C.** Yes, if the lot size were measured in thousands of square feet, the estimate coefficient would be 4 instead of 0.004, thus making th
- **D.** Yes, if the lot size were measured in thousands of square feet, the estimate coefficient would be 1,000 instead of 0.004, thus norm

The degree of freedom to test if the coefficients on *BDR* and *Age* are statistically different from zero at the 1% level is [ ].

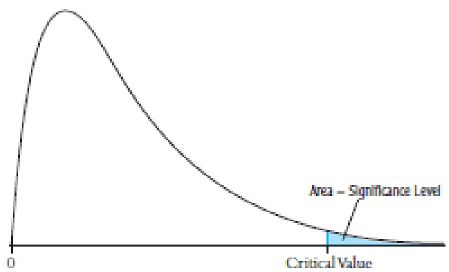
The critical value for the preceding test using the $F_{m,\infty}$ distribution table[1] is [ ]. (*Enter your values exactly as they appear in the table.*)

The *F*-statistic for omitting *BDR* and *Age* from the regression is $F = 0.08$. Are the coefficients on *BDR* and *Age* statistically different from zero at the 1% level?

- **A.** Because 0.08 is greater than the critical value, the coefficients are jointly significant at the 1% level.
- **B.** Because 0.08 is greater than the critical value, the coefficients are not jointly significant at the 1% level.
- **C.** Because 0.08 is less than the critical value, the coefficients are jointly significant at the 1% level.
- **D.** Because 0.08 is less than the critical value, the coefficients are not jointly significant at the 1% level.

1: Critical Values for the F Distribution

Critical Values for the $F_{m,\infty}$ Distribution



Degrees of Freedom	Significance Level 10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64
8	1.67	1.94	2.51
9	1.63	1.88	2.41
10	1.60	1.83	2.32

Answers 0.206

C. Since the *t*-statistic < 1.96, the coefficient on *BDR* is not statistically significantly different from zero.

B.
No, the coefficient on *BDR* measures the partial effect of the number of bedrooms, holding house size constant, and thus significantly underestimates the price of five-bedroom houses.

6.18

9.98

C.
Yes, if the lot size were measured in thousands of square feet, the estimate coefficient would be 4 instead of 0.004, thus making the regression results easy to read and interpret.

2

4.61

D. Because 0.08 is less than the critical value, the coefficients are not jointly significant at the 1% level.

ID: Exercise 7.7

---

11. When testing a joint hypothesis, you should:

- **A.** use the *F*-statistics and reject at least one of the hypotheses if the statistic exceeds the critical value.
- **B.** use *t*-statistics for each hypothesis and reject the null hypothesis once the statistic exceeds the critical value for a single hypothesi
- **C.** use *t*-statistics for each hypothesis and reject the null hypothesis if all of the restrictions fail.
- **D.** use the *F*-statistic and reject all the hypotheses if the statistic exceeds the critical value.

Answer: A. use the *F*-statistics and reject at least one of the hypotheses if the statistic exceeds the critical value.

ID: Test A Ex 7.2.3

---

12. You have estimated the relationship between test scores and the student-teacher ratio under the assumption of homoskedasticity of the error terms. The regression output is as follows: $\widehat{Test\ Score} = 698.9 - 2.28 \times STR$, and the standard error on the slope is 0.48.

The homoskedasticity-only "overall" regression F-statistic for the hypothesis that the regression R^2 is zero is approximately:

- **A.** 4.75.
- **B.** 22.56.
- **C.** 0.96.
- **D.** 1.96.

Answer: B. 22.56.

ID: Test B Ex 7.2.2

---

13. The homoskedasticity-only F-statistic and the heteroskedasticity-robust F-statistic typically are:

- **A.** different.
- **B.** a multiple of each other (the heteroskedasticity-robust F-statistic is 1.96 times the homoskedasticity-only F-statistic).
- **C.** related by a linear function.
- **D.** the same.

Answer: A. different.

ID: Test B Ex 7.2.3

---

14. The critical value of $F_{4,\infty}$ at the 5% significance level is:

- **A.** 3.84.
- **B.** 2.37.
- **C.** 1.94.
- **D.** Cannot be calculated because in practice you will not have an infinite number of observations.

Answer: B. 2.37.

ID: Test B Ex 7.2.4

---

15. Which of the following statements are true in describing the Bonferroni method of testing hypotheses on multiple coefficients? (*Check all that apply*.)

- [ ] **A.** It modifies the "one-at-a-time" method so that it uses different critical values that ensure that its size equals its significance level.
- [ ] **B.** Its advantage is that it applies very generally.
- [ ] **C.** Its advantage is that it can have a very high power and is used especially when the regressors are highly correlated.
- [ ] **D.** It modifies the "one-at-a-time" method by using the *F*-statistic to test joint hypotheses.

Suppose a researcher studying the factors affecting the monthly rent of a one-bedroom apartment (measured in dollars) estimates the following regression using data collected from 150 houses:

$$\textit{Rent} = 548.65 - 2.09\ \textit{Location} + 2.12\ \textit{Neighborhood} - 1.05\ \textit{Crime},$$

where *Location* denotes the distance of the apartment from downtown (measured in miles), *Neighborhood* denotes the average monthly income of the people living in the neighborhood of the apartment, and *Crime* denotes the crime rate within the 5 km radius of the apartment.

The researcher wants to test the hypothesis that the coefficient on *Location*, β_1 and the coefficient on *Neighborhood*, β_2 are jointly zero, against the hypothesis that at least one of these coefficients is nonzero. The test statistics for testing the null hypotheses that $\beta_1 = 0$ and $\beta_2 = 0$ are calculated to be 1.56 and 1.23, respectively. Suppose that these test statistics are uncorrelated.

The *F*-statistic associated with the above test will be ______.

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

At the 5% significance level, we will (1) ______ the null hypothesis.

(1) ○ fail to reject
○ reject

Answers A.
It modifies the "one-at-a-time" method so that it uses different critical values that ensure that its size equals its significance level.
, B. Its advantage is that it applies very generally.

1.97

(1) fail to reject

ID: Concept Exercise 7.2.1

16. Suppose that a researcher selects a random sample of 200 columnists from a large newspaper company to study the factors affecting the productivity of these columnists (measured by the number of words they write in a day). She estimates the following regression equation:

$$\widehat{W} = 632.78 - 0.84\ S + 0.24\ Inc + 1.76\ Exp + 0.84\ HS,$$

where W denotes the number of words they write in a day, S denotes the number of minutes they spend browsing social networking sites in a day, *Inc* denotes the monthly salary they earn, *Exp* denotes the number of years of experience they have, and *HS* denotes their daily overall health measured by a health score on a scale of 1 to 100 which includes various health indicators.

The researcher hypothesizes that after controlling for the social media browsing time and the overall health, neither income nor experience have a significant effect on the productivity of the columnists, i.e., β_2 and β_3 are jointly zero.

The researcher calculates the test statistics for individually testing the null hypotheses $\beta_2 = 0$ and $\beta_3 = 0$ to be 2.55 and 1.46, respectively. Suppose that the correlation between these test statistics is found to be -0.48.

The F-statistic associated with the above test will be ______.

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

At the 1% significance level, we will (1) ______ the null hypothesis.

(1) ○ reject
○ fail to reject

Answers 7.93

(1) reject

ID: Concept Exercise 7.2.2

---

17. A researcher is interested in finding out the factors which determined the yearly spending on family outings last year (Y, measured in dollars). She compiles data on the number of members in a family (X_1), the annual income of the family (X_2), and the number of times the family went out on an outing in the last year (X_3). She collects data from 152 families and estimates the following regression:

$$\hat{Y} = 120.45 + 1.54X_1 + 1.98X_2 + 1.98X_3.$$

Suppose $\beta_1, \beta_2, \beta_3$, denote the population slope coefficients of X_1, X_2, and X_3, respectively.

The researcher wants to check if neither X_1 nor X_2 have a significant effect on Y or at least one of them has a significant effect, keeping X_3 constant. She calculates the value of the F-statistic for the test with the two restrictions (H_0: $\beta_1 = 0$, $\beta_2 = 0$ vs. H_1: $\beta_1 \neq 0$ and/or $\beta_2 \neq 0$) to be 4.61.

The p-value for the test will be ______.

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

At the 5% significance level, we will (1) ______ the null hypothesis.

(1) ○ reject
○ fail to reject

Answers 0.01

(1) reject

ID: Concept Exercise 7.2.3

18. Suppose Coach Jackson wants to study the performance (Y, measured in m/s) of track athletes. He chooses the amount of calorie intake (X_1) and the hours of sleep (X_2) as regressors. He collects data from 150 randomly selected athletes at his academy and estimates the following regression function:

$$\hat{Y} = 4.89 + 0.96X_1 + 0.88X_2.$$

Suppose β_1 denotes the population slope coefficient of X_1.

Coach Jackson wants to test whether or not the number of hours of sleep has a significant effect on an athlete's performance, keeping the amount of calorie intake constant. He calculates the value of the t-statistic for the test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ to be 3.17.

The F-statistic for this test will be ______.

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

At the 5% significance level, the value of the F-statistic suggests that we will (1) ______ the null hypothesis.

(1) ○ reject
○ fail to reject

Answers 10.05

(1) reject

ID: Concept Exercise 7.2.4

19. If the error term is homoskedastic, the F-statistic can be written in terms of the improvement in the fit of the regression. How can we measure this improvement in fit? (*Check all that apply.*)

- [ ] **A.** The improvement in the fit of the regression can be measured by the increase in sum of squared residuals (SSR).
- [ ] **B.** The improvement in the fit of the regression can be measured by the decrease in sum of squared residuals (SSR).
- [ ] **C.** The improvement in the fit of the regression can be measured by the increase in the regression R^2.
- [ ] **D.** The improvement in the fit of the regression can be measured by the decrease in the regression R^2.

A researcher wants to study the factors which affected the sales of cars by different manufacturers in the automobile industry across the world in the year 2017. Generally, the sales of cars (S, measured in thousands) depend on the average price of the cars sold by the manufacturer (P, measured in thousand dollars), the average interest rate at which car loans were offered in that country in that year (I, expressed as a percentage), and the manufacturers' total expenditure on the advertisement of their cars (E, measured in thousand dollars). She selects a random sample of 100 car manufacturers and estimates the following regression function:

$$\hat{S} = 245.73 - 0.75I - 0.45P + 0.75E,$$

By imposing restrictions on the true coefficients, the researcher wishes to test the null hypothesis that the coefficients on I and E are jointly 0, against the alternative that atleast one of them is not equal to 0, while controlling for the other variables.

The values of the sum of squared residuals (SSR) from the unrestricted and restricted regressions are 36.50 and 38.75, respectively.

The homoskedasticity-only F-statistic value associated with the above test will be ______.

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

At the 1% significance level, the researcher will (1) ______ the joint null hypothesis.

(1) ○ reject
○ fail to reject

Answers B. The improvement in the fit of the regression can be measured by the decrease in sum of squared residuals (SSR)., C. The improvement in the fit of the regression can be measured by the increase in the regression R^2.

2.96

(1) fail to reject

ID: Concept Exercise 7.2.5

20. Which of the following statements is true?

- **A.** In an unrestricted regression, the null hypothesis is forced to be true.
- **B.** We would fail to reject the null hypothesis if the sum of squared residuals (*SSR*) from the restricted regression is sufficiently small
- **C.** In a restricted regression, the alternative hypothesis is allowed to be true.
- **D.** We would reject the null hypothesis if the sum of squared residuals (*SSR*) from the unrestricted regression is sufficiently smaller th

A statistics student wants to study the factors which affected the sale of Ben & Jerry's ice creams (S) across the world on last year's National Ice Cream Day. He selects three factors - the average price of the ice creams sold in that region (P), the average temperature on that day in that region (T), and the regional expenditure on advertising their ice cream in the week leading to that day (E). For his study, he selects a random sample of 100 stores and estimates the following regression function:

$$\hat{S} = 2.25 - 0.57P + 0.65T + 0.70E,\ R^2 = 0.45.$$

By imposing restrictions on the true coefficients, the student wishes to test the null hypothesis that the coefficients on T and E are jointly 0 against the alternative that atleast one of them is not equal to 0, while controlling for the other variables.

So, the restricted regression equation is:

$$\hat{S} = 2.25 - 0.57P,\ R^2 = 0.37.$$

The homoskedasticity-only F-statistic value associated with the above test is ______.

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

At the 5% significance level, the student will (1) ____________ the joint null hypothesis.

(1) ○ reject
○ fail to reject

Answers D.

We would reject the null hypothesis if the sum of squared residuals (*SSR*) from the unrestricted regression is sufficiently smaller than that from the restricted regression.

6.98

(1) reject

ID: Concept Exercise 7.2.6

21. Consider the following multiple regression model

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

Which of the following describes how to test the null hypotheses that either $\beta_1 = 0$ or $\beta_2 = 0$?

- **A.** Compute the standard errors, the *t*-statistic and the *p*-value for each, β_1 and β_2. Reject the null hypothesis if the *p*-value is less th
- **B.** Compute the standard errors, the correlation between β_1 and β_2, the *F*-statistic, and the *p*-value associated with the *F*-statistic. Re
- **C.** We fail to reject the null hypothesis only if both estimated coefficient, $\hat{\beta}_1$ and $\hat{\beta}_2$, are precisely equal to zero.
- **D.** We fail to reject the null hypothesis only if either estimated coefficient, $\hat{\beta}_1$ or $\hat{\beta}_2$, is precisely equal to zero.

Which of the following describes how to test the null hypothesis that $\beta_1 = 0$ and $\beta_2 = 0$?

- **A.** Compute the standard errors, the *t*-statistic and the *p*-value for each, β_1 and β_2. Reject the null hypothesis if the *p*-value is less th
- **B.** Compute the standard errors, the correlation between β_1 and β_2, the *F*-statistic, and the *p*-value associated with the *F*-statistic. Re
- **C.** We fail to reject the null hypothesis only if either estimated coefficient, $\hat{\beta}_1$ or $\hat{\beta}_2$, is precisely equal to zero.
- **D.** We fail to reject the null hypothesis only if both estimated coefficient, $\hat{\beta}_1$ and $\hat{\beta}_2$, are precisely equal to zero.

Suppose you want to test the null hypothesis that $\beta_1 = 0$ and $\beta_2 = 0$. Is the result of the joint test implied by the result of the two separate tests?

- **A.** Yes.
- **B.** No.

Answers A.

Compute the standard errors, the *t*-statistic and the *p*-value for each, β_1 and β_2. Reject the null hypothesis if the *p*-value is less than some relevant significance level.

B.

Compute the standard errors, the correlation between β_1 and β_2, the *F*-statistic, and the *p*-value associated with the *F*-statistic. Reject the null hypothesis if the *p*-value is less than some relevant significance level.

B. No.

ID: Review Concept 7.1

22. 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. West is the omitted region. A detailed description of the variables used in the data set is available here .

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 (0.26)	5.39 (0.26)	5.38 (0.26)
Female (X_2)	−2.48 (0.27)	−2.22 (0.27)	−2.22 (0.27)
Age (X_3)		0.22 (0.04)	0.22 (0.04)
Northeast (X_4)			0.56 (0.28)
Midwest (X_5)			0.67 (0.23)
South (X_6)			−0.25 (0.29)
Intercept	12.95 (0.13)	4.67 (1.03)	3.22 (1.08)
Summary Statistics and Joint Tests			
F-statistic for regional effects = 0			6.14
SER	6.27	6.22	6.21
$\bar{R}^2$	0.104	0.168	0.154
n	3365	3365	3365

Note: The numbers in parentheses below each estimated coefficient are the estimated standard errors.

The regression shown in column (2) was estimated again, this time using data from 1992 (3365 observations selected at random from the March 1993 CPS, converted into 1998 dollars using the consumer price index). The results are

$$\widehat{AHE} = \underset{(0.98)}{0.77} + \underset{(0.20)}{5.29}College - \underset{(0.18)}{2.59}Female + \underset{(0.03)}{0.40}Age,\ SER = 5.85,\ \bar{R}^2 = 0.21$$

The numbers in parentheses are the estimated standard errors.

Calculate the *t*-statistic for the change in the *College* coefficient between 1992 and 1998.

The *t*-statistic for the change in the *College* coefficient between 1992 and 1998 ______.

(*Round your response to three decimal places*)

Was the change in the *College* coefficient between 1992 and 1998 statistically significant at the 5% significance level?

- **A.** Yes.
- **B.** No.

Answers 0.305

B. No.

ID: Exercise 7.5

23. Refer to the 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. West is the omitted region. A detailed description of the variables used in the data set is available here .

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)	5.45 (0.24)	5.35 (0.24)	5.49 (0.24)
Female (X_2)	−2.51 (0.21)	−2.84 (0.21)	−2.84 (0.21)
Age (X_3)		0.14 (0.04)	0.14 (0.04)
Northeast (X_4)			0.64 (0.25)
Midwest (X_5)			0.51 (0.22)
South (X_6)			−0.26 (0.26)
Intercept	12.43 (0.13)	4.72 (1.06)	3.15 (1.02)
Summary Statistics and Joint Tests			
F-statistic for regional effects = 0			6.13
SER	6.27	6.22	6.21
R^2	0.152	0.129	0.111
n	3001	3001	3001

Note: The numbers in parentheses below each estimated coefficient are the estimated standard errors.

Evaluate the following statement: "In all of the regressions, the coefficient on *Female* is negative, large, and statistically significant. This provides strong statistical evidence of gender discrimination in the U.S. labor market."

Hint: Consider two identical workers that differ only in gender, and think about the causal relationship between earnings and gender.

- **A.** True.
- **B.** False.

Answer: B. False.

ID: Exercise 7.6

24. Consider the following multiple regression

$$\widehat{Price} = \underset{(23.3)}{118.6} + \underset{(2.66)}{0.578}BDR + \underset{(7.56)}{23.7}Bath + \underset{(0.019)}{0.146}Hsize + \underset{(0.00059)}{0.003}Lsize + \underset{(0.348)}{0.101}Age - \underset{(10.3)}{47.1}Poor,\ \bar{R}^2 = 0.72,\ \text{SER} = 40.1$$

The numbers in parentheses below each estimated coefficient are the estimated standard errors. A detailed description of the variables used in the data set is available here .

Suppose you wanted to test the hypothesis that *BDR* equals zero. That is,

$$H_0: BDR = 0 \text{ vs } H_1: BDR \neq 0$$

Report the *t*-statistic for this test.

The *t*-statistic is [ ]

(*Round your response to three decimal places*)

Is the coefficient on *BDR* statistically different from zero at the 5% significance level?

- A. Yes.
- B. No.

Typically five-bedroom houses sell for much more than two-bedroom houses. Is this consistent with your previous answer and with the regression more generally?

- A. Yes.
- B. No.

A homeowner purchases 2179 square feet from and adjacent lot. Construct a 95% confidence interval for the change in the value of her house.

The 95% confidence interval for the change in the value of the home is [[ ] , [ ]]

(*Round your response to two decimal places*)

Lot size is measured in square feet. Do you think that measuring lot size in thousands of square feet might be more appropriate?

- A. Yes, because small differences in square footage between two houses is not likely to have a significant effect on differences in hou
- B. No, because small differences in square footage between two houses likely have a significant effect on differences in house prices
- C. No, because changing the units in which lot size is measured will likely render the estimated coefficient insignificant.
- D. Yes, because changing the units in which lot size is measured will likely make the estimated coefficient more significant.

The *F*-statistic from the joint test of *BDR* and *Age* is *F* = 0.11. Are the coefficients on *BDR* and *Age* statistically different from zero at the 10% level?

- A. Yes.
- B. No.

Answers 0.217

B. No.

A. Yes.

4.02

9.06

A.
Yes, because small differences in square footage between two houses is not likely to have a significant effect on differences in house prices.

B. No.

ID: Exercise 7.7

25. Refer to the table of estimated regressions below, computed using data for 1999 from all 420 K–6 and K–8 districts in California, to answer the following question. The variable of interest, *test scores*, is the average of the reading and math scores on the Stanford 9 Achievement Test, a standardized test administered to fifth–grade students. School characteristics (average across the district) include enrollment, number of teachers (measured as "full–time equivalents"), number of computers per classroom, and expenditure per student.

Results of Regressions of test scores on the Student-Teacher Ratio and Student Characteristic Control Variables Using California Elementary School Districts.

Dependent variable: average test score in the district.

Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Student–teacher ratio (X_1)	−2.74** (0.51)	−1.85* (0.47)	−1.97** (0.29)	−1.15** (0.31)	−1.38** (0.23)
Percent English learners (X_2)		−0.604** (0.037)	−0.107** (0.034)	−0.458** (0.038)	−0.121** (0.033)
Percent elegible for subsidized lunch (X_3)			−0.564** (0.024)		−0.509** (0.039)
Percent on public income assistance (X_4)				−0.756** (0.063)	0.049 (0.058)
Intercept	694.7** (10.3)	687.1** (8.7)	701.8** (5.4)	701.8** (6.8)	703.7** (5.9)
Summary Statistics and Joint Tests					
SER	18.43	14.23	9.51	11.25	9.44
$\bar{R}^2$	0.045	0.417	0.763	0.628	0.791
n	456	456	456	456	456

These regressions were estimated using data on K-8 school districts in California. Heteroskedastic–robust standard errors are given in parentheses under coefficients. The individual coefficient is statistically significant at the *5% level or **1% significance level using a two–sided test.

Compute the R^2 for each of the regressions.

1. The R^2 for the regression in column (1) is: ______
2. The R^2 for the regression in column (2) is: ______
3. The R^2 for the regression in column (3) is: ______
4. The R^2 for the regression in column (4) is: ______
5. The R^2 for the regression in column (5) is: ______

(*Round your response to three decimal places*)

Construct the homoskedasticity-only *F*-statistic for testing $\beta_3 = \beta_4 = 0$ in the regression shown in column (5).

The homoskedasticity-only *F*-statistic for the test is: ______

(*Round your response to two decimal places*)

Is the homoskedasticity-only *F*-statistic significant at the 5% level?

- **A.** Yes.
- **B.** No.

Test $\beta_3 = \beta_4 = 0$ in the regression shown in column (5) using the Bonferroni test. Note that the 1% *Bonferroni* critical value is 2.807.

The *t*-statistic for β_3 in the regression in column (5) is: ______

(*Round your response to three decimal places*)

The *t*-statistic for β_4 in the regression in column (5) is: ______

(*Round your response to three decimal places*)

Is the Bonferroni test significant at the 1% level?

- A. Yes.
- B. No.

Construct a 99% confidence interval for β_1 for the regression in column (5).

The 99% confidence interval is: [,]

(*Round your response to three decimal places*)

Answers 0.047

0.420

0.765

0.630

0.793

406.34

A. Yes.

− 13.051

0.845

B. No.

− 1.973

− 0.787

ID: Exercise 7.8

26. Consider the two variable regression model:

$$Y_i = \beta_0 + \beta_1 Edu_{1i} + \beta_2 Exp_{2i} + u_i,$$

where Y denotes the average monthly income, *Edu* denotes the number of years of education, *Exp* denotes the number of years of experience, and u_i denotes the error term.

Suppose the researcher wants to test whether the effect of education on average monthly income and the effect of experience on the average monthly income of an individual are the same or not. So, the test the researcher wants to conduct is $H_0: \beta_1 = \beta_2$ vs. $H_1: \beta_1 \neq \beta_2$. The hypotheses can be tested by modifying the original regression equation to turn the restriction into a restriction on a single regression coefficient.

Suppose the regression function is modified in the following way:

$$Y_i = \beta_0 + \gamma_1 Edu_{1i} + \beta_2 W_{1i} + u_i, \text{ where } \gamma_1 = \beta_1 - \beta_2 \text{ and } W_i = Edu_{1i} + Exp_{2i}\ .$$

Since $\gamma_1 = \beta_1 - \beta_2$, the test the researcher wants to conduct will now be $H_0: \gamma = 0$ vs. $H_1: \gamma \neq 0$. Let $\hat{\gamma}_1$ and $SE(\hat{\gamma}_1)$, denote the estimated slope coefficient of γ_1 and the standard error of $\hat{\gamma}_1$, respectively.

If $SE(\hat{\gamma}_1)$ is 0.75 and $\hat{\gamma}_1$ is 1.25, then the 95% confidence interval for the difference between the coefficients, $\beta_1 - \beta_2$, denoted by γ_1, is (_____ , _____).

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

Based on the calculated confidence interval, we can say that at the 5% significance level, we will (1) __________ the null hypothesis $H_0: \gamma = 0$.

(1) ○ fail to reject
○ reject

Answers – 0.22

2.72

(1) fail to reject

ID: Concept Exercise 7.3.1

---

27. When there are two coefficients, the resulting confidence sets are:

○ **A.** rectangles.

○ **B.** trapezoids.

○ **C.** squares.

○ **D.** ellipses.

Answer: D. ellipses.

ID: Test B Ex 7.4.5

---

28. Consider the two variable regression model:

$$Y_i = \beta_0 + \beta_1 Edu_{1i} + \beta_2 Exp_{2i} + u_i,$$

where Y denotes the average monthly income, Edu denotes the number of years of education, Exp denotes the number of years of experience, and u_i denotes the error term.

Suppose the researcher wants to test whether the effect of education on average monthly income and the effect of experience on the average monthly income of an individual are the same or not. So, the test the researcher wants to conduct is H_0: $\beta_1 = \beta_2$ vs. H_1: $\beta_1 \neq \beta_2$.

Which of the following is the modified regression so that hypothesis testing can be carried out using the t-statistic?

- **A.** $Y_i = \beta_0 + \gamma_1 Edu_{1i} + \beta_2 W_{1i} + u_i$, where $\gamma_1 = \beta_1 + \beta_2$ and $W_i = Edu_{1i} - Exp_{2i}$.
- **B.** $Y_i = \beta_0 + \gamma_1 Edu_{1i} + \beta_2 W_{1i} + u_i$, where $\gamma_1 = \beta_1 - \beta_2$ and $W_i = Edu_{1i} + Exp_{2i}$.
- **C.** $Y_i = \beta_0 + \gamma_i Edu_{1i} + \beta_2 W_i + u_i$, where $\gamma_i = Edu_{1i} + Exp_{2i}$ and $W_i = \beta_1 - \beta_2$.
- **D.** $Y_i = \beta_0 + \gamma_i Edu_{1i} + \beta_2 W_i + u_i$, where $\gamma_i = Edu_{1i} - Exp_{2i}$ and $W_i = \beta_1 - \beta_2$.

Suppose the population regression is of the form:

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

The 95% confidence set for the coefficients β_1 and β_2 will be (1) ____________ which contains the pairs of values of β_1 and β_2 that (2) ____________ rejected using the F-statistic at the 5% significance level.

(1)
- an ellipse
- a rectangle
- a circle

(2)
- will be
- cannot be

Answers B. $Y_i = \beta_0 + \gamma_1 Edu_{1i} + \beta_2 W_{1i} + u_i$, where $\gamma_1 = \beta_1 - \beta_2$ and $W_i = Edu_{1i} + Exp_{2i}$.

(1) an ellipse

(2) cannot be

ID: Concept Exercise 7.4.1

29. Suppose you run a regression of test scores against parking lot area per pupil. Is the R^2 likely to be high or low?

- **A.** Low, Because the relationship between test scores and parking lot area per pupil is a causal relationship.
- **B.** High, because parking lot area is correlated with student–teacher ratio, with whether the school is in a suburb or a city, and possib
- **C.** High, Because the relationship between test scores and parking lot area per pupil is a causal relationship.
- **D.** Low, because parking lot area is correlated with student–teacher ratio, with whether the school is in a suburb or a city, and possibl

Are the OLS estimators likely to be biased and inconsistent?

- **A.** The OLS estimators are likely to be unbiased and consistent because the R^2 is probably sufficiently high.
- **B.** The OLS estimators are likely biased and inconsistent because there are omitted variables correlated with parking lot area per pup
- **C.** The OLS estimators are likely to be biased and inconsistent because the R^2 is probably not sufficiently high.
- **D.** The OLS estimators are likely unbiased and consistent because there are no omitted variables correlated with parking lot area per

Answers B.

High, because parking lot area is correlated with student–teacher ratio, with whether the school is in a suburb or a city, and possibly with district income.

B.

The OLS estimators are likely biased and inconsistent because there are omitted variables correlated with parking lot area per pupil that also explain test scores, such as ability.

ID: Review Concept 7.2

---

30. All of the following are true with the exception of one condition:

- **A.** a high R^2 or $\bar{R}^2$ always means that an added variable is statistically significant.
- **B.** a high R^2 or $\bar{R}^2$ does not mean that the regressors are a true cause of the dependent variable.
- **C.** a high R^2 or $\bar{R}^2$ does not necessarily mean that you have the most appropriate set of regressors.
- **D.** a high R^2 or $\bar{R}^2$ does not mean that there is no omitted variable bias.

Answer: A. a high R^2 or $\bar{R}^2$ always means that an added variable is statistically significant.

ID: Test A Ex 7.5.4

---

31. Consider a regression with two variables, in which X_{1i} is the variable of interest and X_{2i} is the control variable.

Conditional mean independence requires:

- **A.** $E\left(\mu_i \middle| X_{1i}, X_{2i}\right) = E\left(\mu_i \middle| X_{2i}\right)$.
- **B.** $E\left(\mu_i\right) = E\left(\mu_i \middle| X_{2i}\right)$.
- **C.** $E\left(\mu_i \middle| X_{1i}\right) = E\left(\mu_i \middle| X_{2i}\right)$.
- **D.** $E\left(\mu_i \middle| X_{1i}, X_{2i}\right) = E\left(\mu_i \middle| X_{1i}\right)$.

Answer: A. $E\left(\mu_i \middle| X_{1i}, X_{2i}\right) = E\left(\mu_i \middle| X_{2i}\right)$.

ID: Test A Ex 7.5.5

32. Which of the following statements best describe how the R^2 and the adjusted R^2 ($\bar{R}^2$) should be interpreted in practice? (*Check all that apply.*)

- [ ] **A.** A high R^2 or $\bar{R}^2$ suggests that the regressors are a true cause of the dependent variable.
- [ ] **B.** An increase in the R^2 or $\bar{R}^2$ does not necessarily mean that an added variable is statistically significant.
- [ ] **C.** A high R^2 or $\bar{R}^2$ suggests that you have the most appropriate set of regressors.
- [ ] **D.** A high R^2 or $\bar{R}^2$ does not mean that there is no omitted variable bias.

Suppose that a set of control variables do not satisfy the conditional mean independence condition, $E(u_i|X_i, W_i) = E(u_i|W_i)$, where X_i denotes the variable or variables of interest and W_i denotes one or more control variables.

If the OLS estimators are jointly normally distributed and each $\hat{\beta}_j$ is distributed $N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right)$, $j = 0,..., k,$ then which of the following statements holds true in this case?

- **A.** The variables of interest fail to explain any variation in the regression model.
- **B.** The control variables explain all the variation in the regression model.
- **C.** There will be no remaining omitted determinants of Y that are correlated with X, even after holding W constant, freeing the model f
- **D.** There will remain omitted determinants of Y that are correlated with X, even after holding W constant, and the result is omitted var

Suppose you have developed and estimated a base specification and a list of alternative specifications. Which of the following statements are true? (*Check all that apply.*)

- [ ] **A.** If the estimates of the coefficients of interest change substantially across specifications, this provides evidence that the original sp
- [ ] **B.** If the estimates of the coefficients of interest change substantially across specifications, this provides evidence that the original sp
- [ ] **C.** If the estimates of the coefficients of interest are numerically similar across the alternative specifications, then this provides eviden
- [ ] **D.** If the estimates of the coefficients of interest are numerically similar across the alternative specifications, then this provides eviden

Answers B. An increase in the R^2 or $\bar{R}^2$ does not necessarily mean that an added variable is statistically significant., D. A high R^2 or $\bar{R}^2$ does not mean that there is no omitted variable bias.

D.
There will remain omitted determinants of Y that are correlated with X, even after holding W constant, and the result is omitted variable bias.

A.
If the estimates of the coefficients of interest change substantially across specifications, this provides evidence that the original specification had omitted variable bias and so might your alternative specifications.
, C.
If the estimates of the coefficients of interest are numerically similar across the alternative specifications, then this provides evidence that the estimates from your base specification are reliable.

ID: Concept Exercise 7.5.1

33. Using the Excel data set, CPS08[2], described[3] in Empirical Exercise 4.1, run a regression of average hourly earnings (*AHE*) on *age* and answer the following questions.

The coefficient on *age* shows that

- **A.** *AHE* increase by \$0.0605 for every one-year increase in age
- **B.** *AHE* increases by \$6.05 for every one-year increase in age
- **C.** *AHE* increase by \$0.605 for every one-year increase in age
- **D.** Age has no association with *AHE*

Now run a regression of average hourly earnings (*AHE*) on *bachelor*, *female* and *age*. Comparing the coefficient on *age* in part (1) with coefficient on *age* when *bachelor* and *female* are included, you could conclude that

- **A.** the coefficient on age in part (1) suffered from omitted variable bias
- **B.** there is no evidence of omitted variable bias because the R^2's in the two regressions are roughly the same
- **C.** There is no evidence of omitted variable bias in the simple regression of AHE on age
- **D.** the coefficient on age is now statistically insignificant from zero with the inclusion of the two other regressors

What is the difference in the expected hourly earnings of a 25-year old male with a college degree as compared to a 30-year old female with a college degree?

- **A.** \$0.74
- **B.** \$7.40
- **C.** \$22.08
- **D.** \$21.34

Given the following hypothesis: H_0: β_{female} =0.0 adjusted for age and education we would

- **A.** Not reject H_0 because even though the coefficient on females is -3.66 is it not statistically significant at the 5% level.
- **B.** Reject H_0 because the t-ratio is 0.211
- **C.** Reject H_0 because the 95% confidence interval does not include zero
- **D.** Not Reject H_0 because the t-statistic associated with the null is 0.211

Consider the regression in part (2). Based on a joint test of the hypothesis H_0: $\beta_{female} = \beta_{bachelor} = 0$ we would

- **A.** We would not reject H_0: because the prob-value is 0.13
- **B.** We would not reject H_0: because the R^2 is less than 0.200, indicative of low explanatory power
- **C.** Reject H_0: because the F^{act} is 822 which is much larger than the critical $F_{2,8}$ of 3.0
- **D.** We would reject H_0: because the F-statistic for the model is 641.4

The adjusted and unadjusted R^2 from the regression in part (2) are very similar.because

- **A.** the regressors have relatively little explanatory power
- **B.** because (n-1)/(n-k-1) is close to 1.0
- **C.** the model F-statistic is so large
- **D.** both equal ESS/TSS

A person with a college degree earns on average \$8.03 more than someone with less than a college degree holding age and gender constant. If the t-statistic and standard error were remove from the output could you still test the following null:
H_0: $\beta_{bachelor} = 4.0$?

- A. No, there would not be enough information
- B. You could use the 95% confidence interval from which you would reject the null
- C. You could use the 95% confidence interval from which you would NOT reject the null
- D. You could use the model F-statistic from which you would reject the null

Rerun the regression in part (2), but drop the variable, *bachelor*. What happens to the coefficient on *female*?

- A. it falls by $1.13 in absolute value which suggests that there is an omitted variable bias when bachelor is excluded
- B. it falls but the difference is trivial
- C. it becomes larger (less negative) suggesting that we increase the bias by including *bachelor* in the model
- D. it becomes more negative suggesting that fewer women get a bachelor's degree relative to men.

Based on the regression in part (8) (AHE on *age* and *female*), test the following: H_0: $\beta_{age} = \beta_{female} = 0$.

- A. there is insufficient information to answer the question.
- B. the R^2 is 0.044 suggesting that you cannot reject the null
- C. the F^{act} statistic is 178.4 so you would reject the null at the 1% level
- D. the prob-value for the model F-statistic is very small indicating that you cannot reject the null

2: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/cps08.xlsx

3: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/CPS08_Description.pdf

Answers C. *AHE* increase by $0.605 for every one-year increase in age

C. There is no evidence of omitted variable bias in the simple regression of AHE on age

A. $0.74

C. Reject H_0 because the 95% confidence interval does not include zero

C. Reject H_0: because the F^{act} is 822 which is much larger than the critical $F_{2,8}$ of 3.0

B. because (n-1)/(n-k-1) is close to 1.0

B. You could use the 95% confidence interval from which you would reject the null

A. it falls by $1.13 in absolute value which suggests that there is an omitted variable bias when bachelor is excluded

C. the F^{act} statistic is 178.4 so you would reject the null at the 1% level

ID: General Empirical 7.1 (static)

34. Using the Excel data set, College Distance[4], described[5] in Empirical Exercise 4.3, run a regression of years of completed schooling (*ed*) on distance (*dist*) from a 4-year college in 10s of miles.

An advocacy group claims that a person's educational attainment would increase by 0.37 years if distance to the nearest college was decreased by 50 miles.

- **A.** True, because the R^2 multiplied by 50 is approximately 0.37
- **B.** False, decreasing distance would decrease schooling by 0.37
- **C.** True, because you multiply the coefficient on distance by 5
- **D.** False, the coefficient on dist is only -0.073

Regress completed schooling (*ed*) on the variables *dist*, *female, black, Hispanic,bytest, dadcoll incomehi, ownhome, cue80,* and *stwmfg80*. With the additional variables the coefficient on distance is

- **A.** no longer statistically significant at the 5% level
- **B.** falls to -0.032 which is too small to be of importance
- **C.** the prob-value for the coefficient on distance is 0.011 indicating that we would not reject the null: H0:â$_{dist}$ = 0.
- **D.** falls to -0.032 but is still statistically significant because the t-statistic is greater than 1.96 in absolute value.

Based on the regression in part (2), the expected value of completed schooling is 14.97 years for a black female with a base year test score of 50, a father that went to college, who is from a family with income greater than $25,000 and that owns its home and who lives in a county were the unemployment rate is 6.0, the state hourly wage in manufacturing is $8.00 and who lives 10 miles from the nearest 4-year college. If distance were increased from 10 to 100 miles and all other characteristics were the same, the expected value of completed schooling would

- **A.** decrease by 0.10
- **B.** remain unchanged
- **C.** decrease to 14.68
- **D.** increase to 15.25

Regress *ed* on *dist, black* and *Hispanic*. The coefficients on *black* and *Hispanic* indicate that both groups

- **A.** obtain less schooling than whites controlling for *dist*
- **B.** obtain more schooling that whites
- **C.** that blacks obtain 0.56 years less schooling than whites
- **D.** both (a) and (c)

Using the results in part (4), test the following, H_0: $\beta_{black} = \beta_{Hispanic} = 0$.

- **A.** Do not reject H_0: because the prob-value for the F statistic is so small
- **B.** Do not reject H_0: because the R^2 is only 0.0218, which is very small
- **C.** Reject H_0: because the F^{act} is 27.8 with 2, and 3,972 df
- **D.** Reject H_0: because the F^{act} is 28.2 with 3, and 3,972 df

Compare the coefficients on *black* and *Hispanic* in part (4) with those in part (2), the full model. The results from the full model indicate that

- **A.** blacks and Hispanics obtain more schooling than whites when adjusted for more than just distance
- **B.** race/ethnicity is no longer an important determinant of schooling
- **C.** blacks and Hispanics obtain less schooling than whites regardless of the variables included in the model
- **D.** adding the other variables did not change the R^2

The model in part (2) has 7 more variables than the model in part (4): *female,bytest, dadcoll incomehi, ownhome, cue80,* and *stwmfg80*. Based on a test of whether these additional variables add statistically significant explanatory power to the model we would conclude that

- **A.** they do not because the adjusted R^2 and partial R^2 are so similar
- **B.** you cannot test it with the information given
- **C.** they do because the F-stat of the model is 142.3 and the prob-value is very small
- **D.** they do because the partial F-statistic is 192.7 which is large relative to the critical F

The coefficient on *cue80* in the regression in part (2) should be interpreted as follows:

- **A.** every 10 percentage points increase in *cue80* increases *ed* by 0.230 years
- **B.** every one percent increase in *cue80* increases *ed* by 0.023 years
- **C.** every one percentage point increase in *cue80* increases *ed* by 0.023 years
- **D.** Both (a) and (c)

Instead of expressing *cue80* as a percent, divide it by 100 and express it as a fraction of the population unemployed. Then re-run the regression in part (2) but use the fraction unemployed instead of the percent unemployed. The coefficient on the fraction unemployed indic that

- **A.** every .01 increase in the fraction unemployed, increases *ed* by 2.3 years
- **B.** every 1.00 increase in the fraction unemployed, increases *ed* by 2.3 years
- **C.** every .01 increase in the fraction unemployed, increases *ed* by 0.023 years
- **D.** Both (b) and (c)

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

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

Answers C. True, because you multiply the coefficient on distance by 5

D. falls to -0.032 but is still statistically significant because the t-statistic is greater than 1.96 in absolute value.

C. decrease to 14.68

D. both (a) and (c)

C. Reject H_0: because the F^{act} is 27.8 with 2, and 3,972 df

A. blacks and Hispanics obtain more schooling than whites when adjusted for more than just distance

D. they do because the partial F-statistic is 192.7 which is large relative to the critical F

D. Both (a) and (c)

D. Both (b) and (c)

ID: General Empirical 7.2 (static)

35. 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), *CCTV* denotes the number of CCTV cameras installed in the district, and *Prison* denotes the number of prison inmates. The following table shows the results of a few regressions of the total cost to the state.

Dependent variable: total cost to the state (in thousand dollars)				
Regressor	**(1)**	**(2)**	**(3)**	**(4)**
LEOP	12.32 (0.52)	17.99 (0.84)	14.55 (2.25)	18.1 (0.82)
DTP		0.73 (0.06)	0.59 (0.12)	0.75 (0.04)
CCTV				0.73 (0.13)
Prison		2.12 (0.5)		2.11 (0.39)
Intercept	182.5 (11.52)	191.6 (6.68)	219.95 (5.26)	288.5 (4.14)
$\bar{R}^2$	0.12	0.75	0.64	0.75
n	220	220	220	220

Heteroskedasticity-robust standard errors are given in parentheses under coefficients.

Which of the following statements correctly describe the reasons behind the differences observed in the coefficients in the given specifications? (*Check all that apply.*)

☐ **A.** According to the 4th specification, reducing the number of law enforcement officers on patrol by one officer is estimated to decrease total cost to the state by approximately $18.10, holding constant other factors.

☐ **B.** The significant rise in the coefficient on *LEOP* from the 1st specification to the 4th shows the presence of omitted variable bias in the 1st specification.

☐ **C.** The value of $\bar{R}^2$ in the 1st specification suggests that the number of law enforcement officers on patrol alone explains a large fraction of the variation in total cost to the state.

☐ **D.** The number of *CCTV* cameras installed in the district appears to be redundant. As reported in regression (4), adding it to regression (2) has a negligible effect on the estimated coefficients on *LEOP* and *DTP* or their standard errors.

Answer: B.

The significant rise in the coefficient on *LEOP* from the 1st specification to the 4th shows the presence of omitted variable bias in the 1st specification.

, D.

The number of *CCTV* cameras installed in the district appears to be redundant. As reported in regression (4), adding it to regression (2) has a negligible effect on the estimated coefficients on *LEOP* and *DTP* or their standard errors.

ID: Concept Exercise 7.6.1