Student: ____________________ **Date:** ____________________	**Instructor:** Richeng Piao **Course:** ECON 2560 - Applied Econometrics	**Assignment:** Practice Problem Set 8

1. The interpretation of the slope coefficient in the model $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \mu_i$ is as follows:

   - **A.** a change in X by one unit is associated with a β_1 change in Y.
   - **B.** a 1% change in X is associated with a β_1% change in Y.
   - **C.** a change in X by one unit is associated with a 100 β_1% change in Y.
   - **D.** a 1% change in X is associated with a change in Y of 0.01 β_1.

   Answer: B. a 1% change in X is associated with a β_1% change in Y.

   ID: Test A Ex 8.1.1

---

2. Assume that you had estimated the following quadratic regression model:

$$\widehat{Test\ Score} = 607.3 + 3.85 Income - 0.0423 Income^2$$

   If income increased from 10 to 11 ($10,000 to $11,000), then the predicted effect on test scores would be:

   - **A.** 3.85 − 0.0423.
   - **B.** 3.85.
   - **C.** 2.96.
   - **D.** Cannot be calculated because the function is nonlinear.

   Answer: C. 2.96.

   ID: Test A Ex 8.1.2

---

3. A nonlinear function:

   - **A.** can be adequately described by a straight line between the dependent variable and one of the explanatory variables.
   - **B.** makes little sense, because variables in the real world are related linearly.
   - **C.** is a concept that only applies to the case of a single or two explanatory variables since you cannot draw a line in four dimensions.
   - **D.** is a function with a slope that is not constant.

   Answer: D. is a function with a slope that is not constant.

   ID: Test B Ex 8.1.1

---

4. An independent researcher in district W wants to study the monthly electricity consumption of households (E, measured in kilowatt hours) in the district. For his study, he selects a random sample of 175 households from the district and estimates the following regression function:

$$\widehat{E} = 102.5 + 3.25N - 0.0195N^2,$$
$$(1.5561)\ (0.6954)\ (0.0061)$$

where N denotes the number of members in the household and standard errors appear in parentheses.

The researcher wants to test the hypothesis that the relationship between number of members in the house and the electricity consumption by households is linear, against the alternative that it is nonlinear.

Suppose β_2 denotes the population slope coefficient on the regressor N^2. The researcher conducts the test $H_0: \beta_2 = 0$ vs. $H_1: \beta_2 \neq 0$.

The test statistic associated with the test the researcher wants to conduct is ______.

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

At the 5% significance level, the researcher will (1) ______ the hypothesis that the relationship between number of members in the house and the electricity consumption by households is linear.

(1) ○ fail to reject
○ reject

Answers −3.20

(1) reject

ID: Concept Exercise 8.1.1

5. A market analyst is interested in estimating the relationship between the miles per gallon of fuel for motorcycles (*Mileage*, measured in mpg) and the age of that vehicle (*Age*, measured in years). She collects information on 400 motorcycles across the automobile industry and estimates the following regression equation:

$$\widehat{Mileage} = \underset{(1.23)}{45.45} - \underset{(1.25)}{2.86}Age - \underset{(2.01)}{0.01}Age^2, \ \bar{R}^2 = 0.541.$$

The standard errors are given in parentheses. She wants to find the effect on mileage of a change in age (ΔAge) of motorcycles.

If the age of a motorcycle increases from 1 to 2 years, the predicted change in $\widehat{Mileage}$ will be [ ] mpg.

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

If the age of a motorcycle increases from 7 to 8 years, the predicted change in its mpg ($\Delta Mileage$) will be [ ] mpg.

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

So, when the age of a motorcycle increases from 7 to 8 years, the decrease in total mileage will be (1) ________ as compared to the change when age increases from 1 to 2 years.

The predicted change in mileage when the age of a motorcycle increases from 5 to 6 years is:

$$\Delta Mileage = \left(\hat{\beta}_0 + \hat{\beta}_1 \times 6 + \hat{\beta}_2 \times 6^2\right) - \left(\hat{\beta}_0 + \hat{\beta}_1 \times 5 + \hat{\beta}_2 \times 5^2\right) = (6-5)\hat{\beta}_1 + (36-25)\hat{\beta}_2 = \hat{\beta}_1 + 11\hat{\beta}_2.$$

If the *F*-statistic testing the hypothesis $\hat{\beta}_1 + 11\hat{\beta}_2 = 0$ is $F = 200.18$, then the 95% confidence interval for the change in the predicted value of mileage is ([ ], [ ]).

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

We reject or fail to reject a hypothesis using the criterion does the hypothesized value fall in our 95% confidence interval. Based on the calculated confidence interval, we can say that at the 5% significance level, we will (2) ________ the hypothesis $\hat{\beta}_1 + 11\hat{\beta}_2 = 0$.

(1) ○ higher
○ lower

(2) ○ fail to reject
○ reject

Answers – 2.89

– 3.01

(1) higher

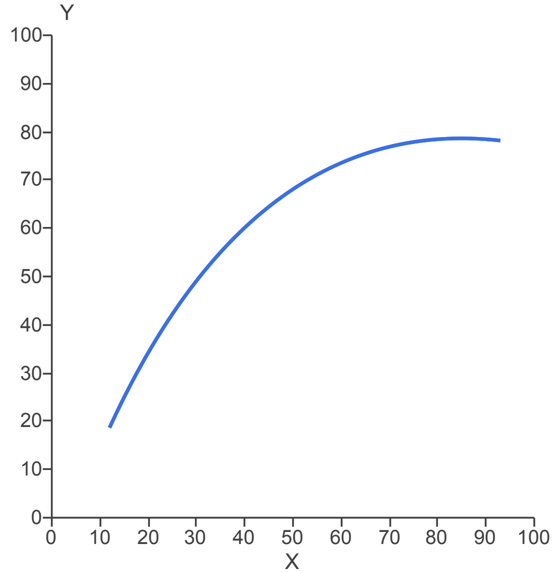
– 3.38

– 2.56

(2) reject

ID: Concept Exercise 8.1.2

6. Consider the following regression function to answer the questions below.



Which of the following specifies a nonlinear regression that model this shape?

- **A.** $Y_i = \exp\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i\right)$.
- **B.** $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$.
- **C.** $Y_i = \beta_0$.
- **D.** $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$.

Which of the following economic relationships may exhibit a shape like this? (*Check all that apply*)

- **A.** The relationship between wage earnings per hour and wage earnings per year.
- **B.** The relationship between wage earnings and years of experience.
- **C.** The relationship between time spent studying for an exam and grade for such exam.
- **D.** The relationship between income and fertility.

Answers B. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$.

B. The relationship between wage earnings and years of experience. , C.
The relationship between time spent studying for an exam and grade for such exam. , D.
The relationship between income and fertility.

ID: Review Concept 8.1

7. A "Cobb-Douglas" production function relates production (Q) to factors of production, capital (K), labor (L), raw materials (M), and an error term u using the equation

$$Q = \lambda K^{\beta_1} L^{\beta_2} M^{\beta_3} e^{u},$$

where λ, β_1, β_2, and β_3 are production parameters.

Suppose that you have data on production and the factors of production from a random sample of firms with the same Cobb-Douglas production function. Which of the following regression functions provides the most useful transformation to estimate the model?

- **A.** A logarithmic regression function.
- **B.** A linear regression function.
- **C.** A quadratic regression function.
- **D.** An exponential regression function.

Answer: A. A logarithmic regression function.

ID: Review Concept 8.2

---

8. A standard "money demand" function used by macroeconomists has the form

$$\ln(m) = \beta_0 + \beta_1 \ln(GDP) + \beta_2 R,$$

Where m is the quantity of (real) money, GDP is the value of (real) gross domestic product, and R is the value of the nominal interest rate measured in percent per year. Supposed that $\beta_1 = 3.29$ and $\beta_2 = -0.04$.

What is the expected change in m if GDP increases by 8%?

The value of m is expected to (1) ________ by approximately [ ] %.

(*Round your response to the nearest integer*)

What is the expected change in m if the interest rate increases from 1% to 6%?

The value of m is expected to (2) ________ by approximately [ ] %.

(*Round your response to the nearest integer*)

(1) ○ increase ○ decrease

(2) ○ increase ○ decrease

Answers (1) increase

26

(2) decrease

20

ID: Review Concept 8.3

---

9. You have estimated a linear regression model relating Y to X. Your professor says, "I think that the relationship between Y and X is nonlinear." How would you test the adequacy of your linear regression? (*Check all that apply*)

- ☐ **A.** Compare the fit between of linear regression to the non-linear regression model.
- ☐ **B.** If adding a quadratic term, you could test the hypothesis that the estimated coefficient of the quadratic term is significantly different from zero.
- ☐ **C.** There is evidence in favor of a non-linear relationship if there is zero correlation between the dependent and independent variable.
- ☐ **D.** All of the above.

Answer: A. Compare the fit between of linear regression to the non-linear regression model. , B. If adding a quadratic term, you could test the hypothesis that the estimated coefficient of the quadratic term is significantly different from zero.

ID: Review Concept 8.4

---

10. A "Cobb-Douglas" production function relates production (Q) to factors of production, capital (K), labor (L), raw materials (M), and an error term u using the equation

$$Q = \lambda K^{\beta_1} L^{\beta_2} M^{\beta_3} e^{u},$$

where λ, β_1, β_2, and β_3 are production parameters. Taking logarithms of both sides of the equation yields

$$\ln(Q) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + u.$$

Suppose that you thought that the value of β_2 was not constant, but rather increased when K increased. Which of the following regression functions captures this dynamic relationship?

- ○ **A.** $\ln(Q) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + \beta_4 [\ln(L) \times \ln(K) \times \ln(M)] + u.$
- ○ **B.** $\ln(Q) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + \beta_4 [\ln(L) \times \ln(K)] + u.$
- ○ **C.** $\ln(Q) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(L)^2 + \beta_4 \ln(M) + u.$
- ○ **D.** $\ln(Q) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(K)^2 + \beta_3 \ln(L) + \beta_4 \ln(M) + u.$

Answer: B. $\ln(Q) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(L) + \beta_3 \ln(M) + \beta_4 [\ln(L) \times \ln(K)] + u.$

ID: Review Concept 8.5

---

11. Sales in a company are \$190 million in 2009 and increase \$206 million in 2010. Compute the percentage increase in sales using the usual formula

$$100 \times \frac{(\text{Sales}_{2010} - \text{Sales}_{2009})}{\text{Sales}_{2009}}$$

Compare this value to the approximation

$$100 \times \left[\ln\left(Sales_{2010}\right) - \ln\left(Sales_{2009}\right)\right]$$

$$100 \times \frac{(\text{Sales}_{2010} - \text{Sales}_{2009})}{\text{Sales}_{2009}} = \boxed{}\%$$

$$100 \times \left[\ln\left(Sales_{2010}\right) - \ln\left(Sales_{2009}\right)\right] = \boxed{}\%$$

(*Express your response as a percentage and round to three places*)

Now, assume that sales in a company are \$190 million in 2009 and increase \$256 million in 2010.

$$100 \times \frac{(\text{Sales}_{2010} - \text{Sales}_{2009})}{\text{Sales}_{2009}} = \boxed{}\%$$

$$100 \times \left[\ln\left(Sales_{2010}\right) - \ln\left(Sales_{2009}\right)\right] = \boxed{}\%$$

(*Express your response as a percentage and round to three places*)

The approximation performs (1) ____________ when the change is small. The quality of the approximation (2) ____________ as the percentage change increases.

(1) ○ better
○ worst

(2) ○ deteriorates
○ improves

Answers 8.421

8.085

34.737

29.815

(1) better

(2) deteriorates

ID: Exercise 8.1

12. Consider the following statement to answer the questions below:

"In my experience, student performance depends on class size, in the following way. Students do well when class size is less than 20 students and do very poorly when class size is greater than 25. There are no gains from reducing class size below 20 students, the relationship is constant in the intermediate region between 20 and 25 students, and there is no loss to increasing class size when it is already greater than 25." To capture these threshold effects, define the following binary variables:

$$STR_{small} = 1 \text{ if } STR < 20, \text{ and } STR_{small} = 0 \text{ otherwise;}$$

$$STR_{moderate} = 1 \text{ if } 20 \le STR \le 25, \text{ and } STR_{moderate} = 0 \text{ otherwise; and}$$

$$STR_{large} = 1 \text{ if } STR > 25, \text{ and } STR_{large} = 0 \text{ otherwise.}$$

Consider the regression

$$TestScore_i = \beta_0 + \beta_1 STRsmall_i + \beta_2 STRlarge_i + u_i$$

What expected signs of β_1 and β_2 are consistent with the statement above?

- **A.** $\beta_1 < 0$ and $\beta_2 > 0$.
- **B.** $\beta_1 > 0$ and $\beta_2 > 0$.
- **C.** $\beta_1 < 0$ and $\beta_2 < 0$.
- **D.** $\beta_1 > 0$ and $\beta_2 < 0$.

A researcher tries to estimate the regression $TestScore_i = \beta_0 + \beta_1 STRsmall_i + \beta_2 STRmoderate_i + \beta_3 STRlarge_i + u_i$ and finds that her computer crashes. Why?

- **A.** The error term of this regression is likely heteroskedastic.
- **B.** The relationship between *TestScore* and class size may be logarithmic.
- **C.** There is perfect multicollinearity between the regressors.
- **D.** The relationship between *TestScore* and class size may be quadratic.

Answers D. $\beta_1 > 0$ and $\beta_2 < 0$.

C. There is perfect multicollinearity between the regressors.

ID: Exercise 8.3

13. This problem is inspired by a study of the "gender gap" in earnings in top corporate jobs [Bertrand and Hallock (2001)]. The study compares total compensation among top executives in a large set of U.S. public corporations in the 1990s. (Each year these publicly traded corporations must report total compensation levels for their top five executives.)

Let *Female* be an indicator variable that is equal to 1 for females and 0 for males. A regression of the logarithm of earnings onto *Female* yields

$$\widehat{\ln(Earnings)} = \underset{(0.01)}{6.48} - \underset{(0.05)}{0.44} Female, \quad SER = 2.65$$

The estimated coefficient on *Female* is -0.44. Explain what this value means.

- **A.** *Earnings* for females are, on average, 44% lower than men's.
- **B.** ln(*Earnings*) for females are, on average, 44% lower than men's.
- **C.** ln(*Earnings*) for females are, on average, 0.44 lower than men's.
- **D.** *Earnings* for females are, on average, 0.44% lower than men's.
- **E.** Both A and C are correct.

The *SER* is 2.65. Explain what this value means.

- **A.** The indicator *Female* explains 2.65% of the variation in ln(*Earnings*).
- **B.** ln(*Earnings*) for females are, on average, 2.65 lower than men's.
- **C.** The indicator *Female* explains 265% of the variation in ln(*Earnings*).
- **D.** The error term has a standard deviation of 2.65 (measured in log-points).

Does this regression suggest that female top executives earn less than top male executives?

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

Does this regression suggest that there is gender discrimination?

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

Answers E. Both A and C are correct.

D. The error term has a standard deviation of 2.65 (measured in log-points).

A. Yes.

B. No.

ID: Exercise 8.7

14. A polynomial regression model is specified as:

- **A.** $Y_i = \beta_0 + \beta_1 X_i + \beta_1^2 X_i + \cdots + \beta_1^r X_i + \mu_i.$
- **B.** $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + \mu_i.$
- **C.** $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_2 + \beta_3 (X_{1i} \times X_{2i}) + \mu_i.$
- **D.** $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Y_i^2 + \cdots + \beta_r Y_i^r + \mu_i.$

Answer: B. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + \mu_i.$

ID: Test A Ex 8.2.3

---

15. In the log-log model, the slope coefficient indicates:

- **A.** $\triangle Y / \triangle X.$
- **B.** the effect that a unit change in X has on Y.
- **C.** $\dfrac{\triangle Y}{\triangle X} \times \dfrac{Y}{X}.$
- **D.** the elasticity of Y with respect to X.

Answer: D. the elasticity of Y with respect to X.

ID: Test A Ex 8.2.4

---

16. The best way to interpret polynomial regressions is to:

- **A.** analyze the standard error of estimated effect.
- **B.** plot the estimated regression function and to calculate the estimated effect on Y associated with a change in X for one or more values of X.
- **C.** take a derivative of Y with respect to the relevant X.
- **D.** look at the t-statistics for the relevant coefficients.

Answer: B.
plot the estimated regression function and to calculate the estimated effect on Y associated with a change in X for one or more values of X.

ID: Test B Ex 8.2.2

---

17. In the model $1n(Y_i) = \beta_0 + \beta_1 X_i + \mu_i$, the elasticity of $E(Y \mid X)$ with respect to X is:

- **A.** $\beta_1 X$.
- **B.** $\dfrac{\beta_1 X}{\beta_0 + \beta_1 X}$.
- **C.** β_1.
- **D.** cannot be calculated because the function is nonlinear.

Answer: A. $\beta_1 X$.

ID: Test B Ex 8.2.3

---

18. Consider the polynomial regression model of degree r, $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \bullet\bullet\bullet + \beta_r X_i^r + \mu_i$.

According to the null hypothesis that the regression is linear and the alternative that is a polynomial of degree r corresponds to:

- **A.** $H_0 : \beta_2 = 0, \beta_3 = 0, \ldots, \beta_r = 0$ vs. H_1 : at least one $\beta_j \neq 0, j = 2, \ldots, r$.
- **B.** $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$.
- **C.** $H_0 : \beta_2 = 0, \beta_3 = 0, \ldots, \beta_r = 0$ vs. H_1 : all $\beta_j \neq 0, j = 2, \ldots, r$.
- **D.** $H_0 : \beta_r = 0$ vs. $H_1 : \beta_r \neq 0$.

Answer: A. $H_0 : \beta_2 = 0, \beta_3 = 0, \ldots, \beta_r = 0$ vs. H_1 : at least one $\beta_j \neq 0, j = 2, \ldots, r$.

ID: Test B Ex 8.2.4

---

19. An independent researcher wants to study the factors which affected the prices set by a leading AC manufacturing company, Chester, across the country for the year 2017. She studies whether the effect on price, of the quantity demanded of ACs by the consumers, is significant or not. She selects a random sample of 150 retail outlets of the company and estimates the following regression equation:

$$\underset{}{\widehat{P}} = \underset{(9.76)}{750.50} - \underset{(8.25)}{12.58} ln(Q),$$

where $\widehat{P}$ is the predicted value of price of the ACs and $ln(Q)$ is the logarithm of the quantity demanded of ACs by the consumers. Standard errors are given in parentheses.

The 95% confidence interval for the slope coefficient β_1 will be (____ , ____).

(*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 hypothesis that $\beta_1 = 0$.

(1) ○ fail to reject
○ reject

Answers − 28.75

3.59

(1) fail to reject

ID: Concept Exercise 8.2.1

20. A student wants to study the impact of the number of kilometers (K) run by a car on its resale price (P, measured in dollars). For her study, she selects a random sample of 100 second hand car sellers from her city and estimates the following regression equation:

$$\hat{P} = 62.75 - 250.25 ln(K),$$

where $\hat{P}$ and K denote the predicted value of the resale price of the car and the number of kilometers run by that car, respectively.

Therefore, a 1% increase in the number of kilometers run by a car is associated with (1) ____________ in the resale price of the car by

[ ] (2) ____________.

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

A researcher is interested in finding out the relationship between the price of house (H, measured in hundred dollars) and its distance from the highway (D, measured in kilometers) passing through her district. She estimates the following regression using 300 observations on prices of house and their corresponding distance from the highway:

$$\widehat{ln(H)} = 60.35 - 0.03D,$$

where $\widehat{ln(H)}$ is the predicted value of the logarithm of house prices and D is the value of the distance of the house from the highway.

Therefore, for the above regression function, an increase in distance of the house from the highway by one kilometer is associated with

(3) ____________ in the price of house by [ ] (4) ____________.

A survey was conducted to study the impact of per capita gross national product or GNP (X, measured in thousand dollars) on life expectancy at birth (Y). Data across 150 countries was collected and the following regression was estimated:

$$ln(Y) = 74.25 + 1.19 ln(X).$$

Therefore, for the above regression, a 1% increase in the per capita GNP of the country is associated with (5) ____________ in the life expectancy at birth by [ ] (6) ____________.

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

(1)	(2)	(3)	(4)	(5)	(6)
○ a decrease	○ percent	○ an increase	○ dollars	○ a decrease	○ dollars
○ an increase	○ dollars	○ a decrease	○ percent	○ an increase	○ percent

Answers (1) a decrease

2.50

(2) dollars

(3) a decrease

3

(4) percent

(5) an increase

1.19

(6) percent

ID: Concept Exercise 8.2.2

21. Suppose the population regression is in the form:

$$\ln Y_i = \beta_0 + \beta_1 X_i + u_i,$$

where β_0, β_1 and u_i represent the intercept, the slope coefficient of X and the error term for the i^{th} observation.

What is the expected value of Y_i given X_i ($E\left(Y_i|X_i\right)$)?

- **A.** $E\left(Y_i|X_i\right) = e^{\beta_0 + \beta_1 X_i} E\left(e^{u_i}|X_i\right)$.
- **B.** $E\left(Y_i|X_i\right) = E\left(\ln\left(\beta_0 + \beta_1 X_i\right) + \ln\left(u_i\right)\right)$.
- **C.** $E\left(Y_i|X_i\right) = E\left(e^{\beta_0 + \beta_1 X_i}|X_i\right)$.
- **D.** $E\left(Y_i|X_i\right) = E\left(e^{\beta_0 + \beta_1 X_i}\right)\left(e^{u_i}|X_i\right)$.

This formula illustrates that we cannot obtain an unbiased estimate of Y by taking the (1) ____________ function of $\beta_0 + \beta_1 X_i$.

Two researchers Andy and Simon are interested in understanding the relationship between total expenditure on durable goods (*Exp*, measured in dollars) and the income of a family (*Income*, measured in dollars) in the last month. They collect data across 200 families. They are however not sure which functional form they should choose which will best describe the data. Andy and Simon estimate two specifications each and individually choose the best among their own specifications. They then meet to choose the specification that fits the data best out of their individually chosen specifications.

Following are the regression functions estimated by Andy:

Model 1: $\widehat{Exp} = 607.13 + 3.85 Income - 0.0423 Income^2$, $\bar{R}^2 = 0.554$.

Model 2: $\widehat{Exp} = 600.13 + 5.12 Income - 0.096 Income^2 + 0.00069 Income^3$, $\bar{R}^2 = 0.555$.

Following are the regression functions estimated by Simon:

Model 3: $\widehat{Exp} = 557.81 + 36.12 \ln(Income)$, $\bar{R}^2 = 0.561$.

Model 4: $\widehat{Exp} = 486.17 + 113.12 \ln(Income) - 26.9[\ln(Income)]^2 + 3.06\left[\ln(Income)^3\right]$, $\bar{R}^2 = 0.560$.

Following are the values of the variables which Andy and Simon use for comparison:

In model 2, the *p*-value for testing the significance of the coefficient on $Income^3$ at 5% significance level was found to be 0.043.

In model 4, the *t*-statistic on the coefficient on the cubic term is 0.818. The *F*-statistic testing the joint hypothesis that the true coefficients on the quadratic and cubic term are both zero was found to be 0.44, with a *p*-value of 0.64 at the 5% significance level.

Based on the above information the model that should be choosen is (2) ____________.

(1)
- exponential
- logarithmic

(2)
- model 1
- model 4
- model 2
- model 3

Answers A. $E\left(Y_i|X_i\right) = e^{\beta_0 + \beta_1 X_i} E\left(e^{u_i}|X_i\right)$.

(1) exponential

(2) model 3

ID: Concept Exercise 8.2.3

22. Suppose that a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below.

Dependent variable: ln(*Price*)					
Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Size	0.00052 (0.000041)				
ln(*Size*)		0.77 (0.058)	0.75 (0.092)	0.62 (2.03)	0.736 (0.063)
$\ln(Size)^2$				0.0088 (0.15)	
Bedrooms			0.0042 (0.041)		
Pool	0.087 (0.032)	0.078 (0.039)	0.089 (0.035)	0.091 (0.038)	0.086 (0.037)
View	0.044 (0.034)	0.028 (0.033)	0.028 (0.029)	0.028 (0.031)	0.028 (0.032)
Pool × *View*					0.0025 (0.14)
Condition	0.19 (0.045)	0.19 (0.036)	0.17 (0.039)	0.14 (0.035)	0.15 (0.037)
Intercept	13.73 (0.072)	6.64 (0.42)	6.73 (0.55)	7.12 (7.54)	6.66 (0.45)
Summary Statistics					
SER	0.111	0.104	0.103	0.108	0.105
$\bar{R}^2$	0.72	0.75	0.77	0.73	0.76

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

Using the results in column (1), what is the expected increase in price of building a 500-square-foot addition to a house, holding everything else in the model constant?

The expected increase in price of building a 500-square-foot addition to a house is [ ]%

(*Express your response as a percentage and round to two decimal places*)

Construct a 95% confidence interval for the percentage change in price.

The 95% confidence interval for the percentage change in price is [[ ] , [ ]]%

(*Express your response as a percentage and round to two decimal places*)

Use $\bar{R}^2$ to compare the regressions in columns (1) and (2). Is it better to use *Size* or ln(*Size*) to explain house prices?

- **A.** There is no difference between using *Size* or ln(*Size*)
- **B.** It is better to use ln(*Size*).
- **C.** It is better to use *Size* only if the distribution of house sizes is not normal.
- **D.** It is better to use *Size*.

Using column (2), what is the estimated effect of a pool on price?

The house price is estimated to increase by [ ]%

(*Express your response as a percentage and round to two decimal places*)

Construct a 95% confidence interval for this effect.

The 95% confidence interval for this effect is [[ ] , [ ]]%

(*Express your response as a percentage and round to two decimal places*)

The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom?

The house price is estimated to increase by [   ] %

(*Express your response as a percentage and round to two decimal places*)

Is the coefficient on *bedrooms* in the regression in column (3) statistically significant?

- **A.** The coefficient is not statistically significant.
- **B.** The coefficient is statistically significant at the 5% significance level.
- **C.** The coefficient is statistically significant at the 1% significance level.
- **D.** The coefficient is statistically significant at the 10% significance level.

How would you interpret the coefficient on *bedrooms* in the regression in column (3)?

- **A.** The coefficient measures the effect of an additional bedroom holding the size of the house, and other factors, constant.
- **B.** The coefficient measures the difference in price between houses with many bedrooms and those with just a few.
- **C.** The coefficient is a binary indicator that equals 1 if the house has more than five bedrooms.
- **D.** The coefficient measures the percentage increase in price associated with a one percent increase in the number of bedrooms holding other factors constant.

Is the quadratic term $\ln(Size)^2$ in the regression in column (4) statistically significant?

- **A.** The coefficient is statistically significant at the 5% significance level.
- **B.** The coefficient is statistically significant at the 1% significance level.
- **C.** The coefficient is not statistically significant.
- **D.** The coefficient is statistically significant at the 10% significance level.

Use the regression in column (5) to compute the expected increase in price when a pool is added to a house without a view.

The house price is estimated to increase by [   ] %

(*Express your response as a percentage and round to two decimal places*)

Use the regression in column (5) to compute the expected increase in price when a pool is added to a house with a view.

The house price is estimated to increase by [   ] %

(*Express your response as a percentage and round to two decimal places*)

What is the difference between the effect of a pool on the price of a house with a view and a house without a view?

The difference in the expected percentage change in price is [   ] %

(*Express your response as a percentage and round to two decimal places*)

Is this difference statistically significant?

- **A.** The difference is statistically significant at the 5% significance level.
- **B.** The difference is not statistically significant.
- **C.** The difference is statistically significant at the 1% significance level.
- **D.** The difference is statistically significant at the 10% significance level.

Answers 26.00

21.98

30.02

B. It is better to use ln(*Size*).

7.80

0.16

15.44

0.42

A. The coefficient is not statistically significant.

A. The coefficient measures the effect of an additional bedroom holding the size of the house, and other factors, constant.

C. The coefficient is not statistically significant.

8.60

8.85

0.25

B. The difference is not statistically significant.

ID: Exercise 8.2

23. Read the box "The Return to Education and the Gender Gap."

The Return to Education and the Gender Gap

Dependent variable: logarithm of *Hourly Earnings.*

Regressor	(1)	(2)	(3)	(4)
Years of education	0.1035** (0.0009)	0.1050** (0.0009)	0.1001** (0.0011)	0.1039** (0.0012)
Female		− 0.263** (0.004)	− 0.432** (0.024)	− 0.451** (0.024)
Female × *Years of education*			0.0121** (0.0017)	0.0134** (0.0017)
Potential experience				0.0143** (0.0012)
Potential experience^2				− 0.000217** (0.000025)
Midwest				− 0.095** (0.006)
South				− 0.092** (0.006)
West				− 0.028** (0.007)
Intercept	1.533** (0.012)	1.629** (0.012)	1.697** (0.016)	1.534** (0.023)
$\bar{R}^2$	0.208	0.258	0.258	0.267

The sample size is 52,970 observations for each regression. *Female* is an indicator variable that equals 1 for women and 0 for men. *Midwest*, *South*, and *West* are indicator variables denoting the region of the United States in which the worker lives: For example, *Midwest* equals 1 if the worker lives in the Midwest and equals 0 otherwise (the omitted region is *Northeast*). Standard errors are reported in parentheses below the estimated coefficients. Individual coefficients are statistically significant at the *5% or **1% significance level.

Scenario A

Consider a man with 18 years of education and 3 years of experience who is from a western state. Use the results from column (4) of the table and the method in Key Concept 8.1 to estimate the expected change in the logarithm of average hourly earnings (*AHE*) associated with an additional year of experience.

The expected change in the logarithm of average hourly earnings (*AHE*) associated with an additional year of experience is ____ %. (*Round your response to two decimal places.*)

Scenario B

Consider a man with 18 years of education and 10 years of experience who is from a western state. Use the results from column (4) of the table and the method in Key Concept 8.1 to estimate the expected change in the logarithm of average hourly earnings (*AHE*) associated with an additional year of experience.

The expected change in the logarithm of average hourly earnings (*AHE*) associated with an additional year of experience is ____ %. (*Round your response to two decimal places.*)

Why are the answers to *Scenario A* and *Scenario B* different?

- **A.** The regression is nonlinear in years of education.
- **B.** The regression is nonlinear for regional variables.
- **C.** The regression is nonlinear in experience.
- **D.** All of the above.

The t-statistic for the difference between the effects in *Scenario A* and *Scenario B* is [ ]. (*Round your response to two decima places.*)

Is the difference between the effects in *Scenario A* and *Scenario B* statistically significant at the 5% level?

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

How would you change the regression if you suspected that the effect of experience on earnings was different for men than for women?

- **A.** Include the interaction term $Female \times (Years\ of\ education)^2$.
- **B.** Include the interaction term $Male \times (Years\ of\ education)^2$.
- **C.** Include interaction terms $Male \times Years\ of\ education$ and $Male \times (Years\ of\ education)^2$.
- **D.** Include interaction terms $Female \times Potential\ experience$ and $Female \times (Potential\ experience)^2$.

Answers 1.28

0.97

C. The regression is nonlinear in experience.

8.68

B. Yes.

D. Include interaction terms $Female \times Potential\ experience$ and $Female \times (Potential\ experience)^2$.

ID: Exercise 8.4

---

24. Consider the population regression of log earnings [Y_i, where Y_i = ln(Earnings$_i$)] against two binary variables: whether a worker is married (D_{1i}, where D_{1i} = 1 if the i^{th} person is married) and the worker's gender (D_{2i}, where D_{2i} = 1 if the i^{th} person is female), and the product of the two binary variables $Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 \left(D_{1i} \times D_{2i}\right) + \mu_i$.

The interaction term:

- **A.** does not make sense since it could be zero for married males.
- **B.** indicates the effect of being married on log earnings.
- **C.** allows the population effect on log earnings of being married to depend on gender.
- **D.** cannot be estimated without the presence of a continuous variable.

Answer: C. allows the population effect on log earnings of being married to depend on gender.

ID: Test B Ex 8.3.5

---

25. A study compares the total earnings of senior officials of 120 large corporations in the U.S. Let *Female* be an indicator variable that equals 1 for females and equals 0 for males, and let *Age* be an indicator variable that equals 1 if the age of the person is greater than 45 and equals 0 otherwise.

The estimated regression equation is as follows:

$$\widehat{Earnings} = 2{,}545.78 - 13.57 Female - 25.74 Age - 47.28 Female \times Age,$$

where *Earnings* denotes the yearly earnings of the officials (measured in thousand dollars).

The predicted mean earnings of males below the age of 45 are $_____.

(*Express your answer in dollars.*)

If Sheila, a senior official at a global firm, turns 46 this year, her predicted mean earnings would (1) ________ by $_____ from last year.

(*Express your answer in dollars.*)

(1) ○ increase
○ decrease

Answers 2,545,780

(1) decrease

73,020

ID: Concept Exercise 8.3.1

26. A medical student wants to study the factors affecting people's cholesterol levels (C). She wants to study whether the effect on cholesterol levels of increasing the weight (W, measured in kilograms or kg) of a person depends on whether the person's intake of saturated fats (F, measured in grams) is high or low. For her study, she randomly selects a sample of 150 people and estimates the following regression equation:

$$\widehat{C} = 82.68 + 0.65W + 2.85F + 1.72(W \times F),$$

where the binary variable, F equals 1 if the amount of saturated fats in the body is high (greater than 500 grams) and equals 0 otherwise.

The difference between the effects of weight on the cholesterol level due to high and low levels of saturated fat is ______.

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

A researcher wants to study the factors affecting the tourist industry in his country. To study the factors, he uses the data on the number of tourists (T, measured in thousands) visiting his country, the advertising expenditure (E, measured in thousand dollars) incurred by the government on promoting the tourist place, and the region in which the tourist place is located (R). He randomly selects 100 tourist places in the country and estimates the following regression equation:

$$\widehat{T} = 55.25 + 0.92E + 1.38R,$$

where the region in which the tourist place is located is the binary variable. It equals 1 if the place is located in the tropical region, and equals 0 otherwise.

The researcher wants to tests whether the intercepts of the regression line when $R = 0$ and $R = 1$, are the same, against the alternative that the intercepts are different.

Suppose β_{01} denotes the intercept of the regression line when $R = 0$ and β_{02} denotes the intercept of the regression when $R = 1$. The standard error of the difference between these intercepts, $SE(\beta_{01} - \beta_{02})$, is 0.91.

So, the test statistic associated with the study the researcher wants to conduct is ______.

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

Since the test statistic is (1) ______ than the critical value in this case, we would (2) ______ the null hypothesis.

(1) ○ greater
○ less

(2) ○ reject
○ fail to reject

Answers 1.72

−1.52

(1) less

(2) fail to reject

ID: Concept Exercise 8.3.2

27. 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*, measured in hundred dollars) is decided on the basis of the age of the person (*A*) and the duration for which the insurance is taken (*D*). A health insurance company collects data on 350 randomly selected customers. The estimated regression function is:

$$\underset{}{\widehat{P}_i} = \underset{(0.98)}{178.47} + \underset{(1.11)}{0.97}\, A + \underset{(1.14)}{1.78}\, D + \underset{(1.21)}{0.57}\, A \times D,\ \bar{R}^2 = 0.546.$$

The standard errors are given in parentheses.

The change in the premium due to an increase in the age of a person by 1 year, holding the duration of insurance constant at 3 years, will be $[ ].

The change in the premium due to an increase in duration of the insurance by 1 year, holding the age of the person constant at 50 years, will be $[ ].

(*Express your answers in dollars*.)

Suppose the age of a person is 50 years and the duration for which he purchases the insurance is 3 years.

If both the age of the person and the duration of insurance increase by 1 year, the change in the premium will be $[ ].

(*Express your answer in dollars.*)

Answers 268

3,028

3,353

ID: Concept Exercise 8.3.3

28. A survey was conducted by a researcher to study the impact of per capita gross national product (*GNP*, measured in thousand dollars), female literacy rate (*FemLit*, number of literate females as a percentage of the total female population, expressed in percentage terms - e.g., 10% expressed as 10) on infant mortality rate (*IMR*, number of deaths per 1,000 live births of children under one year of age, expressed in logarithmic terms). Data across 100 countries were collected and the following regression was estimated:

$$\widehat{IMR} = \underset{(1.23)}{13.52} - \underset{(1.47)}{1.01} FemLit - \underset{(1.08)}{0.78} GNP - \underset{(0.79)}{0.64} FemLit \times GNP, \quad \bar{R}^2 = 0.545.$$

The standard errors are given in parentheses. The researcher wants to check if the effect of a unit increase in *FemLit* and *GNP*, above and beyond the sum of the effects of a unit increase in *FemLit* alone and a unit increase in *GNP* alone is significant or not.

The *t*-statistic of the test the researcher wants to conduct keeping other variables constant will be ______.

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

At the 5% significance level, the researcher should (1) ______ the hypothesis that the effect on *IMR* of *FemLit* does not significantly depend on *GNP*.

Suppose the researcher does not include the interaction term $GNP \times FemLit$ into the regression equation. He finds that all the estimated regression coefficients remain the same as in the previous case.

Suppose the values of *GNP* and *FemLit* are \$5.12 thousand and 45%, respectively.

The effect on *IMR* of an increase in *GNP* by \$1,000 in this case would be (2) ______ than the effect on *IMR* of this increase when the interaction term was included in the regression by ______.

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

(1) ○ ~wrrong
○ fail to reject

(2) ○ less
○ greater

Answers −0.81

(1) fail to reject

(2) greater

28.80

ID: Concept Exercise 8.3.4

29. Consider the following regression equation:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \left(X_i \times D_i\right) + u_i,$$

where β_0, β_1, β_2, and u_i are the intercept, the slope coefficient on X_i, the coefficient on the interaction term which is the product of ($X_i \times D_i$), where D_i is the binary variable respectively.

This regression equation has (1) ____________ slope and (2) ____________ intercept for the two values of the binary variable.

In a multiple regression, the interaction term between its two independent variables X_1 and X_2 is their product $X_1 \times X_2$.

The coefficient on ($X_1 \times X_2$) is the effect of a one-unit increase in the product of X_1 and X_2, above and beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.

This holds true (3) ____________ continuous or binary.

Which of the following statements describes a way of determining the degree of the polynomial in X which best models a nonlinear regression? Let r denote the highest power of X that is included in the regression.

- **A.** A way is to check if the coefficients in the regression equation associated with the smallest values of r are equal to zero.
- **B.** A way is to check if the coefficients in the regression equation associated with the largest values of r are equal to zero.
- **C.** A way is to check if the coefficients in the regression equation associated with the smallest values of r are less than zero.
- **D.** A way is to check if the coefficients in the regression equation associated with the largest values of r are less than zero.

(1) the same / a different

(2) the same / a different

(3) whether X_1 and/or X_2 are / only if X_1 and X_2 are / only if either X_1 or X_2 is

Answers (1) a different

(2) the same

(3) whether X_1 and/or X_2 are

B. A way is to check if the coefficients in the regression equation associated with the largest values of r are equal to zero.

ID: Concept Exercise 8.3.5

30. This problem is inspired by a study of the "gender gap" in earnings in top corporate jobs [Bertrand and Hallock (2001)]. The study compares total compensation among top executives in a large set of U.S. public corporations in the 1990s. (Each year these publicly traded corporations must report total compensation levels for their top five executives.)

Let *Female* be an indicator variable that is equal to 1 for females and 0 for males. A regression of the logarithm of earnings onto *Female* yields

$$\widehat{\ln(Earnings)} = \underset{(0.01)}{6.42} - \underset{(0.05)}{0.42}Female, \; SER = 2.54.$$

Calculate the average hourly earnings for top male and female executives.

The hourly earnings for top male executives is $ [ ] per hour. (*Round your response to two decimal places.*)

The hourly earnings for top female executives is $ [ ] per hour. (*Round your response to two decimal places.*)

What is the estimated average difference between earnings of top male executives and top female executives?

The estimated average difference between earnings of top male executives and top female executives is $ [ ] per hour. (*Round your response to two decimal places.*)

What is the estimator of the standard deviation of the regression error?

The estimator of the standard deviation of the regression error is [ ]. (*Round your response to two decimal places.*)

Calculate the *t*-statistic for *Female*.

The *t*-statistic for *Female* is [ ]. (*Round your response to two decimal places.*)

Looking at the *t*-statistic, does this regression suggest that female top executives earn less than top male executives?

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

Does this imply that there is gender discrimination?

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

Two new variables, the market value of the firm (a measure of firm size, in millions of dollars) and stock return (a measure of firm performance, in percentage points), are added to the regression:

$$\widehat{\ln(Earnings)} = \underset{(0.03)}{3.86} - \underset{(0.04)}{0.28}Female + \underset{(0.004)}{0.37}\ln(MarketValue) + \underset{(0.003)}{0.004}Return,$$

$$n = 46{,}670, \; \bar{R}^2 = 0.345.$$

If *MarketValue* increases by 2.48%, what is the increase in earnings?

If *MarketValue* increases by 2.48%, earnings increase by [ ] (1) ________. (*Round your response to two decimal places.*)

The coefficient on *Female* is now − 0.28. Why has it changed from the first regression?

- **A.** *MarketValue* is important for explaining ln(*Earnings*).
- **B.** *Female* is correlated with the two new included variables.
- **C.** The first regression suffered from omitted variable bias.
- **D.** All of the above.

Assume that the coefficient estimated in the second regression is correct. Forget about the effect of the *Return* variable, whose effect seems small and statistically insignificant. Calculate the correlation between *Female* and ln(*MarketValue*) using the omitted variable bias equation.

Let X = *Female*, u = *MarketValue*, and $\frac{\sigma_u}{\sigma_X} = 0.49$.

The correlation between *Female* and ln(*MarketValue*), ρ_{Xu}, is [ ]. (*Round your response to three decimal places.*)

Are large firms more likely to have female top executives than small firms?

- **A.** There is no relationship between the genders.
- **B.** No.
- **C.** Yes.

(1) ○ %
○ dollars per hour

Answers 614.00

403.43

210.57

2.54

– 8.40

A. Yes.

B. No.

0.92

(1) %

D. All of the above.

– 0.286

B. No.

ID: Exercise 8.7

31. Open the Excel data set, CPS08[1], described in Empirical Exercise 4.1. The variables are described[2] in the Word file, CPS_Description.docx. Regress average hourly earns (*AHE*) on *age*, *female* and *bachelor*.

If age increase from 35 to 36, by how much is *AHE* expected to change?

- **A.** \$0.058
- **B.** \$0.585
- **C.** \$0.370
- **D.** \$0.740

Re-run the regression in part (1) but use the natural logarithm of *AHE* as the dependent variable. The effect on an increase in age from 35 to 36 is?

- **A.** 0.027 dollars
- **B.** 0.027 years
- **C.** 2.73 percent
- **D.** 0.273 percent

Re-run the regression in part (2) but use the natural logarithm of age (*ln age*) instead of age. Remember the dependent variable is the *Ln AHE.* The expected change in *AHE* given an increase in age from 35 to 36 is

- **A.** 0.262 percent
- **B.** 2.26 percent
- **C.** 0.026 percent
- **D.** 2.62 dollars

Re-run the regression in part (1) but add the square of age to the model (include both *age* and *age2*). What is the expected *AHE* of a 40-year old woman, with a college degree?

- **A.** \$26.23
- **B.** \$25.51
- **C.** \$22.70
- **D.** \$29.24

Based on your results in part (4) you would conclude that the quadratic in age is

- **A.** an unnecessary addition because the natural logarithm of *age* used in part (2) is superior
- **B.** an appropriate functional form because the sum of the coefficients *age2* and *age2* is larger than the coefficient on *age*when entered alone as in part (1)
- **C.** an appropriate functional form because the coefficient on *age2* is statistically significant at the 5% level
- **D.** an unnecessary addition because the coefficient on *age2* is not statistically significant at the 5% level

Test the following: H_0: $\beta age = \beta age2 = 0$. Based on this test you would

- **A.** Do not reject H_0 because the t-statistic for each coefficient is less than 1.96 in absolute value.
- **B.** Do not reject H_0 because the $F^{act} = 130.9$ is smaller than the critical of $F_{2,7706}$
- **C.** Reject H_0 because the $F^{act} = 29.1$ which is larger than the critical of $F_{2,7706}$
- **D.** Reject H_0 because the $F^{act} = 130.9$ which is larger than the critical of $F_{2,7706}$

Age is a rough proxy for experience. It has been proposed that women of the same age and maybe the same experience as men earn les: every year they age. Regress *Ln AHE* on *bachelor*, *female*, *age* and the interaction of *age* and *female*. Based on this regression you conclu that

- **A.** Women's wages increase 3.46% less than men's wages for every increment in age
- **B.** Women's wages increase 3.46% less than men's wages for every increment in age
- **C.** Women's wages increase 1.68% less than men's wages for every increment in age
- **D.** Women's wages increase 30.98% less than men's wages for every increment in age

Based on the results in part (7), there is no statistical support for the proposition that women's wage increase less than men's wages for ea year that they age.

- **A.** False, the coefficient on the interaction term is negative and statistically significant
- **B.** True, the coefficient on the interaction term is statistically insignificant
- **C.** False, the coefficient on age is positive and statistically significant
- **D.** True, the coefficient on the interaction term is irrelevant to the question

The effect of age on *Ln AHE* is different for high school graduates than for college graduates?

- **A.** True, the t-statistic on the interaction term is greater than 1.96
- **B.** True, the prob-value on the interaction term is less than 0.05 so it's statistically significant
- **C.** True, college graduates earn 1.7% more per year of age than do high school graduates
- **D.** All the above

Based on the results in part (9), the expected wage (in logs) of a 40-year old woman with a bachelor's degree is

- **A.** 3.30
- **B.** 3.02
- **C.** 2.63
- **D.** 4.40

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

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

Answers B. $0.585

C. 2.73 percent

B. 2.26 percent

A. $26.23

D. an unnecessary addition because the coefficient on *age2* is not statistically significant at the 5% level

D. Reject H_0 because the F^{act} = 130.9 which is larger than the critical of $F_{2,7706}$

C. Women's wages increase 1.68% less than men's wages for every increment in age

A. False, the coefficient on the interaction term is negative and statistically significant

D. All the above

A. 3.30

ID: Empirical Exercise 8.2 (static)

32. Open the Excel data set, College Distance[3], described in Empirical Exercise 4.3. The variables are described[4] in the Word file, CollegeDistance_DataDescription.docx. Regress years of completed schooling (*ed*) on *dist*, *female*, *bytest*, *tuition*, *black*, *Hispanic*, *incomehi*, *ownhome*, *dadcoll*, *momcoll*,*cue80* and *stwmfg80*.

The effect on an increase in distance from 1 to 2 (10 to 20 miles) would change the expected years of education by how much holding all other factors constant?

- **A.** -0.740
- **B.** -0.074
- **C.** -0.370
- **D.** -0.037

Re-run the regression in part (1) but use the natural logarithm of *ed* (ln *ed*) as the dependent variable. The effect on an increase in distance from 1 to 2 (10 to 20 miles) would change the expected years of education by how much holding all other factors constant?

- **A.** -0.026 years
- **B.** -0.26 percent
- **C.** -0.0026 percent
- **D.** -0.26 years

Re-run the regression in part (1) but add the square of distance to the model (include both *dist* and *dist2*). What is the change in expected *ed* as distance increases from 1 to 2 (10 to 20 miles)?

- **A.** -0.076 years
- **B.** -0.067 years
- **C.** -0.067 percent
- **D.** -0.081 years

Based on your results in part (3) you would conclude that the quadratic in distance is

- **A.** An appropriate functional form because the coefficient on *dist2* is statistically significant at the 5% level
- **B.** an appropriate functional form because the sum of the coefficients *dist* and *dist2* is larger than the coefficient on *dist*when entered alone as in part (1)
- **C.** unnecessary because the coefficient on *dist2* is not statistically significant at the 5% level
- **D.** unnecessary because the natural logarithm of *dist* used in part (2) is superior

Consider 4 boys, Carlos, Tom, Darnell, and Zhen. Assume that their families all have the same values of *dist*, *bytest*, *tuition*, *black*, *Hispanic*, *incomehi*, *ownhome*, *cue80* and *stwmfg80*. Further both of Carlos's parent graduated from college; Tom's father graduated from college but not his mother; Darnell's mother graduate from college but not his father; and neither of Zhen's parents graduated from college. Using the regression results from part (1), answer the following 3 questions.

The expected difference in education between Carlos and Zhen is

- **A.** 0.378 years
- **B.** 0.571 years
- **C.** 0.949 years
- **D.** 0.193 years

The expected difference in education between Carlos and Darnell is

- **A.** 0.378 years
- **B.** 0.571 years
- **C.** 0.193 years
- **D.** 0.949 years

The expected difference in education between Tom and Darnell is

- A. 0.378 years
- B. 0.949 years
- C. 0.193 years
- D. 0.571 years

Suppose you believed that high income families were more able to afford tuition and because of this individuals from high income families h even more education the higher the level of tuition level than families with low income. Create an interaction term that tests this proposition Add it to the model in part (1). Based on this regression and holding all other variables constant, what is the expected difference in years of education for a person from a high-income family facing a tuition of 10K (note tuition is in 10,000s of dollars) relative to a person from a low income family?

- A. 0.365 fewer years
- B. 0.127 fewer years
- C. 0.127 more years
- D. 0.365 more years

Based on the regression results in part (8) you would

- A. Drop the interaction term of high income and tuition, since the coefficient is too small to matter in predictions
- B. Drop the interaction term of high income and tuition since the coefficient is negative which does not make sense
- C. Keep the interaction term of high income and tuition, since the coefficient is statistically significant from zero at the 5 percent level.
- D. Drop the interaction term of high income and tuition since the coefficient is not statistically significant from zero

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 D. -0.037

B. -0.26 percent

B. -0.067 years

A. An appropriate functional form because the coefficient on *dist2* is statistically significant at the 5% level

C. 0.949 years

B. 0.571 years

C. 0.193 years

D. 0.365 more years

D. Drop the interaction term of high income and tuition since the coefficient is not statistically significant from zero

ID: General Empirical 8.1 (static)

33. Consider the following least squares specification between test scores and the student-teacher ratio:

$$\widehat{\textit{Test Score}} = 557.8 + 36.42\ 1\text{n}(\textit{Income})$$

According to this equation, a 1% increase in income is associated with an increase in test scores of:

- **A.** 36.42 points.
- **B.** 557.8 points
- **C.** 0.36 points.
- **D.** cannot be determined from the information given here.

Answer: C. 0.36 points.

ID: Test A Ex 8.4.5

34. An agriculture economist wants to study the factors affecting the production of milk (P, in million liters) across the world in the year 2017. For her study, she selects the number of dairy cows (N), the average age of the cows (A), frequency of milking the cow (F) and the climate of that particular region (C) as regressors. She selects a random sample of 180 countries. The following table shows the results of a few regressions of the production of milk.

Dependent variable: production of milk (in million liters)				
Regressor	**(1)**	**(2)**	**(3)**	**(4)**
***Number of dairy cows* (*N*)**	10.52** (0.52)	15.75** (0.84)	14.55** (2.25)	19.75** (0.82)
***Average age of the* cows (*A*)**		8.25** (0.05)	7.89** (0.06)	8.92** (0.09)
***Climate* (*C*, *binary:Temperate=1, Otherwise = 0*)**				4.65** (3.65)
Frequency of milking the cows (*F*)		5.68** (0.58)	6.75** (0.71)	7.56** (0.65)
F^2			− 3.65** (1.28)	− 4.25** (1.75)
F^3			0.067** (0.035)	0.078** (0.045)
$C \times F$				1.36** (0.82)
$C \times F^2$				− 1.78** (1.36)
$C \times F^3$				0.042** (0.012)
Intercept	95.85 (9.28)	106.17 (10.20)	115.95 (8.26)	123.5 (4.25)
$\bar{R}^2$	0.42	0.70	0.72	0.75
n	180	180	180	180
F-Statistics and p-values on Joint Hypotheses				
***A* and *F* = 0**		4.96 (0.031)		
$C \times F$, $C \times F^2$ and $C \times F^3 = 0$				5.91 (0.001)
F^2 and $F^3 = 0$			5.81 (0.003)	
Standard errors are given in parentheses under coefficients. Individual coefficients are statistically significant at the **1% significance level. *P*-values are given in parentheses under *F*-statistics.				

Which of the following statements are true regarding the specifications given in the above table? (*Check all that apply.*)

- ☐ **A.** The value of $\bar{R}^2$ in the 1st specification suggests that the number of cows alone explains 42% of the variation in the production of milk.
- ☐ **B.** At the 1% significance level, the joint *F*- statistic testing the joint significance of all three interaction terms in the 4th specification − $C \times F$, $C \times F^2$ and $C \times F^3$, suggests that we will fail to reject the null hypothesis.
- ☐ **C.** In the 3rd specification, *F*-statistic testing the hypothesis that the coefficients on F^2 and F^3 are zero is rejected at the 1% significance level.
- ☐ **D.** At the 1% significance level, the *F*- statistic testing the joint significance of *A* and *F* in the 2nd specification suggests that we will reject the null hypothesis.

Answer: A.

The value of $\bar{R}^2$ in the 1st specification suggests that the number of cows alone explains 42% of the variation in the production milk.

, C.

In the 3rd specification, *F*-statistic testing the hypothesis that the coefficients on F^2 and F^3 are zero is rejected at the 1% significance level.

ID: Concept Exercise 8.4.1