Student: ____________________ **Date:** ____________________	**Instructor:** Richeng Piao **Course:** ECON 2560 - Applied Econometrics	**Assignment:** Practice Problem Set 4

1. Binary variables:

   - **A.** can take on only two values.
   - **B.** can take on more than two values.
   - **C.** are generally used to control for outliers in your sample.
   - **D.** exclude certain individuals from your sample.

   Answer: A. can take on only two values.

   ID: Test A Ex 4.1.1

---

2. In which of the following relationships does the intercept have a real-world interpretation?

   - **A.** Weight and height of individuals.
   - **B.** The demand for coffee and its price.
   - **C.** The relationship between the change in the unemployment rate and the growth rate of real GDP ("Okun's Law").
   - **D.** Test scores and class size.

   Answer: C. The relationship between the change in the unemployment rate and the growth rate of real GDP ("Okun's Law").

   ID: Test A Ex 4.1.2

---

3. In the simple linear regression model, the regression slope:

   - **A.** represents the elasticity of Y on X.
   - **B.** when multiplied with the explanatory variable will give you the predicted Y.
   - **C.** indicates by how many percent Y increases, given a one percent increase in X.
   - **D.** indicates by how many units Y increases, given a one-unit increase in X.

   Answer: D. indicates by how many units Y increases, given a one-unit increase in X.

   ID: Test B Ex 4.1.1

---

4. In the simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + \mu_i$:

- **A.** $\beta_0 + \beta_1 X_i$ represents the sample regression function.
- **B.** the intercept is typically small and unimportant.
- **C.** $\beta_0 + \beta_1 X_i$ represents the population regression function.
- **D.** the absolute value of the slope is typically between 0 and 1.

Answer: C. $\beta_0 + \beta_1 X_i$ represents the population regression function.

ID: Test B Ex 4.1.2

---

5. Assume that you have collected a sample of observations from over 100 households and their consumption and income patterns. Using these observations, you estimate the following regression $C_i = \beta_0 + \beta_1 Y_i + \mu_i$, where C is consumption and Y is disposable income.

The estimate of β_1 will tell you:

- **A.** $\dfrac{\triangle \text{Predicted Consumption}}{\triangle \text{Income}}$.
- **B.** $\dfrac{\triangle \text{Income}}{\triangle \text{Predicted Consumption}}$.
- **C.** The amount you need to consume to survive.
- **D.** $\dfrac{\text{Predicted Consumption}}{\text{Income}}$.

Answer: A. $\dfrac{\triangle \text{Predicted Consumption}}{\triangle \text{Income}}$.

ID: Test B Ex 4.1.3

6.

The slope of a regression line is an indicator of the relationship between the (1) ____________.

Suppose that a local law enforcement chief wants to decide whether or not to hire more law enforcement officers to patrol area *X*. For this, the chief wants to know about the current presence of law enforcement officers in area *X* and the reduction in the incidence of crime in that area. In the last 3 years, the total number of law enforcement officers hired for patrolling area *X* increased from 1,350 to 1,820. In the same span, the number of criminal incidents recorded in area *X* decreased from 11,522 to 10,985. The relationship the chief wants to estimate is:

$$\textit{Reported Crimes} = \beta_0 + \beta_{\textit{Officers}} \times \textit{Officers} + \mu_{\textit{other}},$$

where β_0 and $\beta_{\textit{Officers}}$ are the coefficients of the regression line, and $\mu_{\textit{other}}$ is the error term which includes all other factors which could affect the number of reported crimes apart from the presence of law enforcement officers.

From the given information, the distinct effect of changing the number of law enforcement officers on the number of reported crimes, $\beta_{\textit{Officers}}$, is ______.

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

(1)
- ○ intercept and the error term
- ○ dependent and independent variable
- ○ intercept and the independent variable

Answers (1) dependent and independent variable

− 1.14

ID: Concept Exercise 4.1.1

7. Suppose Coach Theil wants to measure the performance of all the track athletes. He wants to specially track the effect of their diet (*Diet*, measured in grams or g) on their field performance (*Performance*, measured in meters per second or m/s). Other coaches also include intensity of the field training (*Training*) and sufficient rest (*Rest*) as parameters while evaluating athletes' performances.

If β_0 denotes the intercept of the population regression line and μ_{other} denotes the error term, which of the following equations describes the regression Coach Theil wants to run?

- **A.** $Performance = \beta_0 + \beta_{Diet} \times Diet + \mu_{other}$.
- **B.** $Diet = \beta_0 + \beta_{Rest} \times Performance + \mu_{other}$.
- **C.** $Performance = \beta_0 + \beta_{Training} \times Training + \mu_{other}$.
- **D.** $Diet = \beta_0 + \beta_{Performance} \times Performance + \mu_{other}$.

The variables explicitly excluded from Coach Theil's population regression line are incorporated in (1) ________.

Which of the following statements are true in describing the coefficients of the regression line when $\beta_0 = 6.2$ m/s? (*Check all that apply*.)

- **A.** For a given diet *X*, the intercept shows the predicted value of performance of the athletes when the diet is zero. As a zero diet is im
- **B.** The slope, β_{Diet}, is the percentage change in performance due to a unit change in the diet.
- **C.** For a given diet *X*, the intercept shows the predicted value of performance of the athletes when the diet is zero. So, the value of inte
- **D.** The slope, β_{Diet}, is the change in performance due to a unit change in the diet.

Suppose the slope of the regression line is 0.10 and no other factors affect athletes' performance apart from their diet.

If the athletes' performance improved by 3.30 m/s within a month, the change in their diet in this month would have been ________ g.

(*Round your answer to one decimal place*.)

(1)
- β_{Diet}
- μ_{other}
- β_{Rest}
- $\beta_{Training}$
- $\beta_{Performance}$
- β_0

Answers A. $Performance = \beta_0 + \beta_{Diet} \times Diet + \mu_{other}$.

(1) μ_{other}

A.
For a given diet *X*, the intercept shows the predicted value of performance of the athletes when the diet is zero. As a zero diet is improbable, the value of the intercept would have no real-world meaning in this case.
, D. The slope, β_{Diet}, is the change in performance due to a unit change in the diet.

33

ID: Concept Exercise 4.1.2

8. Suppose you are interested in studying the relationship between education and wage. More specifically, suppose that you believe the relationship to be captured by the following linear regression model,

$$Wage = \beta_0 + \beta_1 Education + u$$

Suppose further that you estimate the unknown population linear regression model by OLS.

What is the difference between β_1 and $\hat{\beta}_1$?

- **A.** $\hat{\beta}_1$ is a true population parameter, the slope of the population regression line, while β_1 is the OLS estimator of $\hat{\beta}_1$.
- **B.** Both, β_1 and $\hat{\beta}_1$, are OLS estimators of true parameters of the population regression line.
- **C.** β_1 is a true population parameter, the slope of the population regression line, while $\hat{\beta}_1$ is the OLS estimator of β_1.
- **D.** Both, β_1 and $\hat{\beta}_1$, are true parameters of the population regression line.

What is the difference between u and $\hat{u}$?

- **A.** u represents the deviation of observations from the population regression line, while $\hat{u}$ is the difference between *Wage* and its predi
- **B.** u represents the deviation of observations from the population regression line, while $\hat{u}$ is the OLS estimator of *Wage*.
- **C.** u represents the intercept of the population regression line, while $\hat{u}$ is the difference between *Wage* and its predicted value $\widehat{Wage}$.
- **D.** $\hat{u}$ represents the deviation of observations from the population regression line, while u is the difference between *Wage* and its predi

What is the difference between the OLS predicted value $\widehat{Wage}$ and $E(Wage|Education)$?

- **A.** $\widehat{Wage}$ is the expected value of *Wage* for given values of *Education*, while $E(Wage|Education)$ is the OLS predicted value of *Wage* fo
- **B.** $E(Wage|Education)$ is the expected value of *Wage* for given values of *Education*, while $\widehat{Wage}$ is the OLS predicted value of *Wage* fo
- **C.** $E(Wage|Education)$ is the true value of *Wage* for given values of *Education*, while $\widehat{Wage}$ is the OLS predicted value of *Wage* for giv
- **D.** $E(Wage|Education)$ and $\widehat{Wage}$ are equivalent representations of the true value of *Wage* for given values of *Education*.

Answers C. β_1 is a true population parameter, the slope of the population regression line, while $\hat{\beta}_1$ is the OLS estimator of β_1.

A.

u represents the deviation of observations from the population regression line, while $\hat{u}$ is the difference between *Wage* and its predicted value $\widehat{Wage}$.

B.

$E(Wage|Education)$ is the expected value of *Wage* for given values of *Education*, while $\widehat{Wage}$ is the OLS predicted value of *Wage* for given values of *Education*.

ID: Review Concept 4.1

9. Consider the regression model

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

Suppose that you know $\beta_0 = 0$. Derive the formula for the least squares estimator of β_1.

The least squares objective function is

- A. $\sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right)^3$
- B. $\sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right)^2$
- C. $\sum_{i=1}^{n} \left(Y_i^2 - b_0 - b_1 X_i^2\right)$
- D. $\sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right)$

The least squares objective function is $\sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right)^2$. Since we know $\beta_0 = 0$, this is equivalent to $\sum_{i=1}^{n} \left(Y_i - 0 - b_1 X_i\right)^2$. Differentiating with respect to b_1 yields

$$\frac{d \sum_{i=1}^{n} \left(Y_i - 0 - b_1 X_i\right)^2}{db_1} = (1) \_\_\_\_\_\_\_\_$$

Setting this to zero, and solving for the least squares estimator of β_1 yields

$$\widehat{\beta_1} = (2) \_\_\_\_\_\_\_\_$$

(1)

- $-2 \sum_{i=1}^{n} X_i \left(Y_i - 0 - b_1 X_i\right)$
- $-2 \sum_{i=1}^{n} Y_i \left(Y_i - 0 - b_1\right)$
- $-2 \sum_{i=1}^{n} \left(Y_i - 0 - b_1\right)$

(2)

- $\dfrac{\sum_{i=1}^{n} X_i \left(Y_i - 0\right)}{\sum_{i=1}^{n} X_i^2}$
- $\dfrac{\sum_{i=1}^{n} Y_i \left(X_i - 0\right)}{\sum_{i=1}^{n} X_i^2}$
- $\dfrac{\sum_{i=1}^{n} X_i \left(Y_i - 0\right)}{\sum_{i=1}^{n} Y_i^2}$

Answers B. $\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2$

(1) $-2\sum_{i=1}^{n} X_i\left(Y_i - 0 - b_1 X_i\right)$

(2) $\dfrac{\sum_{i=1}^{n} X_i\left(Y_i - 0\right)}{\sum_{i=1}^{n} X_i^2}$

ID: Exercise 4.11

---

10. Consider the regression model

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

The OLS estimators of the slope β_1 and the intercept β_0 are

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

The sample regression line passes through the point $\left(\overline{X}, \overline{Y}\right)$.

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

Answer: B. True

ID: Exercise 4.14

---

11. Changing the units of measurement—that is, measuring test scores in 100s, will do all of the following *except* for changing the:

- **A.** numerical value of the slope estimate.
- **B.** interpretation of the effect that a change in *X* has on the change in *Y*.
- **C.** numerical value of the intercept.
- **D.** residuals.

Answer: B. interpretation of the effect that a change in *X* has on the change in *Y*.

ID: Test A Ex 4.2.3

---

12. The OLS residuals, $\widehat{\mu}_i$, are sample counterparts of the population:

- **A.** regression function intercept.
- **B.** regression function's predicted values.
- **C.** errors.
- **D.** regression function slope.

Answer: C. errors.

ID: Test B Ex 4.2.4

---

13. To decide whether the slope coefficient indicates a "large" effect of X on Y, you look at the:

- **A.** value of the intercept.
- **B.** regression R^2.
- **C.** size of the slope coefficient.
- **D.** economic importance implied by the slope coefficient.

Answer: D. economic importance implied by the slope coefficient.

ID: Test B Ex 4.2.5

---

14. Suppose that you want to estimate the relationship between people's weight (*W*) and the number of times they eat out in a month (*EO*):

$$W_i = \beta_0 + \beta_1 EO_i + u_i,$$

where β_0 is the intercept of the population regression line; β_1 is the slope of the population regression line; u_i is the error term; and the subscript *i* runs over observations, $i = 1, \ldots, n$.

For this, you collect data from a random sample of 250 people. After analyzing the data, you determine that the covariance between people's weight and the number of times they eat out in a month is 4.62 and the variance of the number of times people eat out in a month is 3.78. You also find that the mean weight of people in the sample is 62.36 kg and the mean number of times people eat out in a month is 2.25.

The OLS estimator of the slope β_1 is ____.

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

The OLS estimator of the intercept β_0 is ____.

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

Which of the following statements are true in describing the estimates of the coefficients of the regression line? (*Check all that apply*.)

- ☐ **A.** The intercept shows the predicted weight when an individual does not go out to eat. The value of the intercept therefore has no re
- ☐ **B.** The slope, β_1, is the change in weight due to a unit change in the number of times a person eats out in a month.
- ☐ **C.** The intercept shows the predicted weight when an individual does not go out to eat. The value of the intercept has real-world mea
- ☐ **D.** The slope, β_1, is the percentage change in weight due to a unit change in the number of times a person eats out in a month.

Answers 1.22

59.62

B. The slope, β_1, is the change in weight due to a unit change in the number of times a person eats out in a month. , C. The intercept shows the predicted weight when an individual does not go out to eat. The value of the intercept has real-world meaning in this case.

ID: Concept Exercise 4.2.1

15. Suppose that a random sample of 220 columnists is selected from a large newspaper company and that the number of words they write in a day and the amount of time they spend browsing social networking sites in a day is recorded. A regression of the number of words they write in a day (*W*) on the number of minutes they spend browsing social networking sites in a day (*S*) yields:

$$\widehat{W} = 535.32 - 0.64 \times S.$$

Suppose a columnist spends 88 minutes browsing social networking sites.

The regression's prediction of the number of words the columnist would write that day is ______.

(*Round your answer to the nearest whole number.*)

Suppose the same columnist decides to reduce the time spent on browsing social networking sites by 30 minutes.

On average, the predicted number of words the columnist can write that day would (1) ______ by ______.

(*Round your answer to the nearest whole number.*)

(1) ○ decrease
○ increase

Answers 479

(1) increase

19

ID: Concept Exercise 4.2.2

16. Suppose that a researcher, using data on the number of years of education a person has received and their annual earnings from 150 individuals, estimates the OLS regression:

$$\widehat{Earnings} = 39.74 + 1.58\ Education,$$

where *Earnings* is measured in thousands of dollars.

If the number of years of education a person receives increases by three, on average, the annual earnings of the individual would

(1) ________ by $[ ].

Note: Report answer in dollars.

Suppose that Meghan wants to earn at least $68,000 annually and she has recently earned a Bachelor's degree in Economics, completing 16 years of education in total. She is considering whether she should pursue a Master's degree as well, which would involve studying for two more years.

Which of the following statements is true?

- **A.** She could have reached her income target without a Bachelor's degree by receiving education for only 12 years.
- **B.** She is able to reach her income target with her current level of education.
- **C.** She will not able to reach her income target even after spending two more years earning a Master's degree.
- **D.** She will reach her income target once she spends two more years earning a Master's degree.

(1) ○ increase
○ decrease

Answers (1) increase

4,740

D. She will reach her income target once she spends two more years earning a Master's degree.

ID: Concept Exercise 4.2.3

17. A regression of the gas mileage of a car (*Mileage*, measured in miles per gallon or mpg) on the car's weight (measured in kilograms or kg) using a random sample of cars weighing 900 − 4,200 kg yields the following:

$$\widehat{Mileage} = 55.18 - 0.0082 \times Weight.$$

If the average weight of cars in the sample is 3,580 kg, the average gas mileage in the sample will be ______ mpg.

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

Suppose that a car weighs 3,580 kg and its actual mileage is 25.84 mpg.

The OLS residual in this case is ______ mpg.

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

Suppose you want to use the estimated regression line to predict the mileage of a bus which weighs around 16,000 kg.

Which of the following statements is true?

- **A.** According to the OLS regression line, the predicted mileage given the weight of the bus is 186.38 mpg.
- **B.** According to the OLS regression line, the predicted mileage given the weight of the bus is 25.84 mpg.
- **C.** The estimate from the OLS regression line would be unreliable as the data did not include vehicles weighing more than 4,200 kg.
- **D.** The predicted mileage given the weight of the bus will be the same as the average gas mileage in the sample, i.e., 25.82 mpg.

Answers 25.82

0.02

C.
The estimate from the OLS regression line would be unreliable as the data did not include vehicles weighing more than 4,200 kg.

ID: Concept Exercise 4.2.4

18. Suppose that a researcher, using data on class size (CS) and average test scores from 102 third-grade classes, estimates the OLS regression

$$\widehat{TestScore} = 530.808 + (-5.9364) \times CS, \; R^2 = 0.07, \; SER = 11.7.$$

A classroom has 22 students. The regression's prediction for that classroom's average test score is ______. (*Round your response to two decimal places.*)

Last year a classroom had 19 students, and this year it has 23 students.

The regression's prediction for the change in the classroom average test score is ______. (*Round your response to two decimal places.*)

The sample average class size across the 102 classrooms is 21.83.

The sample average of the test scores across the 102 classrooms is ______. (*Hint:* Review the formulas for the OLS estimators.) (*Round your response to two decimal places.*)

The sample standard deviation of test scores across the 102 classrooms is ______. (*Hint:* Review the formulas for the R^2 and *SER*.) (*Round your response to one decimal place.*)

Answers 400.21

-23.75

401.22

12.1

ID: Exercise 4.1

19. Suppose that a random sample of 200 twenty-year-old men is selected from a population and that these men's height and weight are recorded. A regression of weight on height yields

$$\widehat{Weight} = (-105.3746) + 4.1764 \times Height,\ R^2 = 0.86,\ SER = 10.8,$$

where *Weight* is measured in pounds and *Height* is measured in inches.

What is the regression's weight prediction for someone who is 74 inches tall?

The regression's weight prediction for someone who is 74 inches tall is ______ pounds. (*Round your response to two decimal places.*)

A man has a late growth spurt and grows 1.3 inches over the course of a year. What is the regression's prediction for the increase in this man's weight?

The regression's prediction for the increase in this man's weight is ______ pounds. (*Round your response to two decimal places.*)

Suppose that instead of measuring weight and height in pounds and inches these variables are measured in centimeters and kilograms (1 *in* = 2.54 *cm* and 1 *lb* = 0.4536 *kg*).

Suppose the regression equation in centimeter-kilogram units is:

$$\widehat{Weight} = \hat{\gamma}_0 + \hat{\gamma}_1 Height.$$

The regression estimates from this new centimeter-kilogram regression are:

$\hat{\gamma}_0$ = ______ kg. (*Round your response to four decimal places.*)

$\hat{\gamma}_1$ = ______ kg per cm. (*Round your response to four decimal places.*)

R^2 = ______. (*Round your response to two decimal places.*)

SER = ______ kg. (*Round your response to four decimal places.*)

Answers 203.68

5.43

-47.7979

0.7458

0.86

4.8989

ID: Exercise 4.2

20. A regression of average weekly earnings (*AWE*, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 731.5350 + 10.0800 \times Age,\ R^2 = 0.024,\ SER = 655.3.$$

The coefficient [ ] shows the marginal effect of *Age* on *AWE*; that is, *AWE* is expected to increase by $[ ] for each additional year of age.

[ ] is the intercept of the regression line. It determines the overall level of the line.

(Round your responses to four decimal places.)

The standard error of the regression (*SER*) is 655.3. What are the units of measurement for the *SER*?

- **A.** Dollars.
- **B.** Dollars per week.
- **C.** Dollars per year.
- **D.** Unit-free.

The regression R^2 is 0.024. What are the units of measurement for the R^2?

- **A.** Dollars.
- **B.** Dollars per year.
- **C.** Dollars per week.
- **D.** Unit-free.

What is the regression's predicted earnings for a 25-year-old worker?

The regression's predicted earnings for a 25-year-old worker are $[ ]. (*Round your response to two decimal places.*)

Will the regression give reliable predictions for a 79-year-old worker?

- **A.** No, the oldest worker in the sample is 65 years old; 79 years is far outside the range of the sample data.
- **B.** Yes, although the oldest worker in the sample data is 65 years old, the model is developed to make forecasts and predictions for w

Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal?

- **A.** No, the distribution of earnings is positively skewed and has kurtosis larger than the normal.
- **B.** No, the distribution of earnings is negatively skewed and has kurtosis smaller than the normal.
- **C.** Yes, the distribution of earnings is symmetric and thus normal.

The average age in this sample is 43.7 years. What is the average value of *AWE* in the sample?

The sample mean of *AWE* is $[ ]. (*Round your response to two decimal places.*)

Answers 10.0800

10.0800

731.5350

A. Dollars.

D. Unit-free.

983.54

A. No, the oldest worker in the sample is 65 years old; 79 years is far outside the range of the sample data.

A. No, the distribution of earnings is positively skewed and has kurtosis larger than the normal.

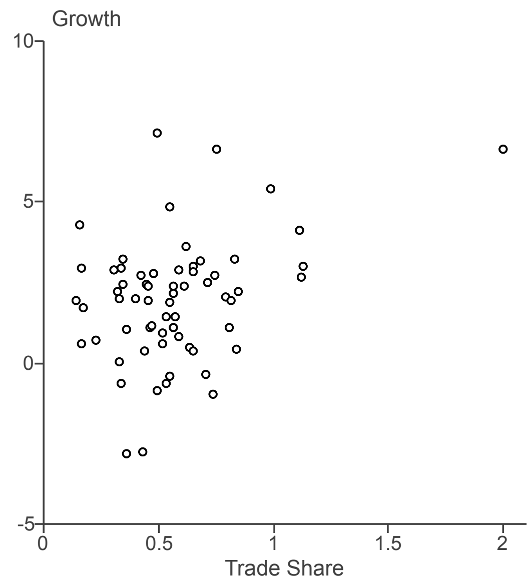
1,172.03

ID: Exercise 4.3

21. In this exercise, you will investigate the relationship between growth and trade.

The following table contains data on average growth rates from 1960 through 1995 for 20 countries along with variables that are potentially related to growth. Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here . Use a statistical package of your choice to answer the following questions.

Construct a scatterplot of average annual growth rate (*Growth*) on the average trade share (*Trade Share*).



Does there appear to be a relationship between the variables?

- **A.** Yes, there appears to be a positive relationship between the variables.
- **B.** There is no evident relationship between the variables.
- **C.** There appears to be a positive relationship between the variables at low levels of a country's trade share, but a negative relationsh
- **D.** Yes, there appears to be a negative relationship between the variables.

One country, Malta, has a trade share (2.000) much larger than the other countries. Should Malta be considered an outlier?

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

Malta is an island nation in the Mediterranean Sea, south of Sicily. Malta is a freight transport site, which explains its large "trade share". Should Malta be included or excluded from the analysis?

- **A.** Malta should be included in the analysis.
- **B.** Malta should be excluded from the analysis.

Suppose you are interested in estimating the following model

$$Growth = \beta_0 + \beta_1 TradeShare + u$$

Run a regression of average annual growth rate (*Growth*) on the country's trade share (*TradeShare*).

What is the estimated intercept $\hat{\beta}_0$?

The estimated intercept $\hat{\beta}_0$ is ______.

(*Round your response to three decimal places*)

What is the estimated slope $\hat{\beta}_1$?

The estimated slope $\hat{\beta}_1$ is ______.

(*Round your response to three decimal places*)

Is the estimated intercept $\hat{\beta}_0$ meaningful in this case?

- A. Yes.
- B. No.

Country A has a trade share of 0.7. Predict Country A's average annual growth rate using the estimated regression.

Country A's predicted annual growth rate is [ ]%.

(*Round your response to two decimal places*)

Country B has a trade share of 0.2. Predict Country B's average annual growth rate using the estimated regression.

Country B's predicted annual growth rate is [ ]%.

(*Round your response to two decimal places*)

Answers A. Yes, there appears to be a positive relationship between the variables.

A. Yes.

B. Malta should be excluded from the analysis.

1.909

0.044

B. No.

1.94

1.92

ID: Empirical Exercise 4.1

22. Open the Excel file, Growth.xls[1], described in Empirical Exercise 4.4. A description[2] of the data is available in the Word file, Growth Description.

Create a scatter plot with growth rates (*Growth*) on the vertical axis and trade shares (*TradeShares*) on the horizontal axis. The relationship between the two appears to be

- **A.** negative
- **B.** no relationship
- **C.** negative but nonlinear
- **D.** positive

Right click any data point and click the option to "Add trend line." In the format option that pops up check the circle that will fit a linear trend and then check the boxes that will display the equation of the line and display the R^2. When that is done, answer the next two questions:

Based on the trend line, the expected growth rate of a country with a trade share of 0.50 would be

- **A.** 1.79
- **B.** 2.95
- **C.** 2.31
- **D.** 0.64

An R^2 of 0.124 indicates that

- **A.** 12.4 percent of the variation in the growth rates can be explained by variation in a country's trade share
- **B.** 1.24 percent of the variation in the growth rates can be explained by variation in a country's trade share
- **C.** 124 percent of the variation in the growth rates can be explained by variation in a country's trade share
- **D.** .124 percent of the variation in the growth rates can be explained by variation in a country's trade share

Use the Excel Regression module (Data/Data Analysis/Regression) and regress the growth rate on trade shares. What is the intercept in the fitted regression line and how do you interpret it?

- **A.** 2.31 and it is the expected growth rate of a county with no trade
- **B.** 0.64 and it is the expected growth rate of a county with no trade
- **C.** None of the above
- **D.** 2.31 and it is the expected growth rate of a county evaluated at the mean of trade shares for the sample of 65 countries

The slope obtained from the regression of growth on trade shares indicates that

- **A.** Every 1 percentage point increase in the trades share decreases growth by 2.31 percent
- **B.** Every 1 percentage point increase in the trades share increase growth by 0.0064 percent
- **C.** Every 1 percentage point increase in the trades share increases growth by 0.023 percent
- **D.** Every 1 percentage point increase in the trades share decreases growth by 0.64 percent

Based on the regression of growth on trade shares, a country that increases its trade share from 0.50 to 0.80 would expect to change its rate of growth by

- **A.** 2.31 percent
- **B.** -0.69 percent
- **C.** None of the above
- **D.** 0.69 percent

Create a scatter plot with growth rates (*Growth*) on the vertical axis and years of schooling for adults (*yearsschool*) on the horizontal axis. Find the data point for The Republic of Korea on the plot. Does this data point make sense or is it an outlier that should be dropped?

- **A.** It makes NO sense since Korea is a well-developed country now and so this must be a mistake
- **B.** It makes NO sense since a country with this low a level of schooling could not grow at 7.2 percent per year for 25 years
- **C.** It makes sense since the data is for schooling in 1960 and the education level of Korean adults has grown rapidly since 1960
- **D.** It makes sense since the data is for schooling in 1960 and the education level of Korean adults has grown steadily at 2.2%

Consider the regression output in part (4). Let G_i be the growth rate for country "i", and let T_i be trade share for that country. Thus the following sum, $\sum_i \left(G_i - \hat{\beta}_0 - \hat{\beta}_1 T_i\right)^2$ is equal to

- **A.** 201.85
- **B.** 1.79
- **C.** 28.49
- **D.** 230.34

In Excel compute the following sums: $\sum_i \left(T_i - \bar{T}\right)^2$, $\sum_i \left(G_i - \bar{G}\right)\left(T_i - \bar{T}\right)$, where T_i is the trade share and G_i is the growth rate in county "
The two sums, respectively, are:

- **A.** 5.36, 12.35
- **B.** 10.50, 15.21
- **C.** 20.81, -29.25
- **D.** 1.91, 0.56

The ratio of $\sum_i \left(G_i - \bar{G}\right)\left(T_i - \bar{T}\right)$, to $\sum_i \left(T_i - \bar{T}\right)^2$ is :

- **A.** 2.31 or the expected growth given a trade share of 0.50
- **B.** 2.31 or an estimate of the intercept in the regression of growth on trade shares
- **C.** 2.31 or the estimated slope in the regression of growth on trade shares
- **D.** 2.31 or an estimate of the covariance between growth and trade shares

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

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

Answers D. positive

A. 1.79

A. 12.4 percent of the variation in the growth rates can be explained by variation in a country's trade share

B. 0.64 and it is the expected growth rate of a county with no trade

C. Every 1 percentage point increase in the trades share increases growth by 0.023 percent

D. 0.69 percent

C. It makes sense since the data is for schooling in 1960 and the education level of Korean adults has grown rapidly since 1960

A. 201.85

A. 5.36, 12.35

C. 2.31 or the estimated slope in the regression of growth on trade shares

ID: Empirical Exercise 4.1 (static)

23. In this exercise, you will investigate the relationship between earnings and height.

These data are taken from the US National Health Interview Survey for 1994. Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here . Use a statistical package of your choice to answer the following questions.

What is the median value of height in the sample?

The median value of height in the sample is ______.

(*Round your response to two decimal places*)

Estimate average earnings for workers whose height is at most 67 inches.

Average earnings for workers whose height is at most 67 inches is estimated to be $______.

(*Round your response to two decimal places*)

Estimate average earnings for workers whose height is greater than 67 inches.

Average earnings for workers whose height is greater than 67 inches is estimated to be $______.

(*Round your response to two decimal places*)

On average, how much more (in absolute value) do taller workers earn compared to shorter workers?

Taller workers earn $______ more than shorter workers on average.

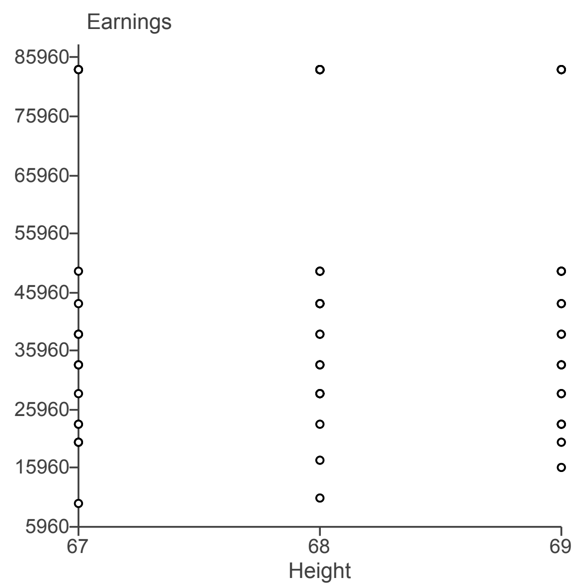
(*Round your response to two decimal places*)

Construct a 95% confidence interval for the absolute value of the difference in average earnings?

The 95% confidence interval for the difference in average earnings is $[______, ______]

(*Round your response to two decimal places*)

Consider the following scatterplot of annual earnings (*Earnings*) on height (*Height*). Why do the points on the plot fall along horizontal lines?



- **A.** Because the variable *Height* is discrete.
- **B.** Because the variable *Height* is continuous.
- **C.** Because the variable *Earnings* is continuous.
- **D.** Because the variable *Earnings* is discrete.

Run a regression of *Earnings* on *Height*.

What is the estimated slope $\hat{\beta}_1$?

The estimated slope $\hat{\beta}_1$ is ______.

(*Round your response to two decimal places*)

Use the estimated regression to predict the earnings for a worker who is 70 inches tall.

Earnings for a worker who is 70 inches tall are predicted to be $ ______.

(*Round your response to two decimal places*)

Suppose height were measured in centimeters instead of inches. Answer the following questions about *Earnings* on *Height* (in cm) regress
Note that 1 inch = 2.54 centimeters.

What is the new estimated slope $\hat{\beta}_1$?

The estimated slope $\hat{\beta}_1$ is ______.

(*Round your response to two decimal places*)

What is the new estimated intercept $\hat{\beta}_0$?

The estimated intercept $\hat{\beta}_0$ is ______.

(*Round your response to two decimal places*)

What is the R^2 of the regression?

The R^2 of the regression is ______.

(*Round your response to four decimal places*)

What is the standard error of the regression?

The standard error of the regression is ______.

(*Round your response up to the nearest integer*)

Do you think that height is uncorrelated with other factors that affect earnings? That is, do you think that the regression error term u_i has a conditional mean of zero given *Height* (X_i)?

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

Answers 68.00

42071.76

57976.42

15904.66

3295.323

28513.997

A. Because the variable *Height* is discrete.

1411.35

57388.57

555.65

−41405.93

0.0013

26875

B. No.

ID: Empirical Exercise 4.2

24. In this exercise, you will investigate the relationship between a worker's age and earnings. (Generally, older workers have more job experience, leading to higher productivity and earnings.)

The following table contains data for full-time, full-years workers, age 25-34, with a high school diploma or B.A./B.S. as their highest degree. Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here . Use a statistical package of your choice to answer the following questions.

Suppose you are interested in estimating the following model

$$Ahe = \beta_0 + \beta_1 Age + u$$

Run a regression of average hourly earnings (*AHE*) on age (*Age*).

What is the estimated intercept $\hat{\beta}_0$?

The estimated intercept $\hat{\beta}_0$ is [ ].

(*Round your response to four decimal places*)

What is the estimated slope $\hat{\beta}_1$?

The estimated slope $\hat{\beta}_1$ is [ ].

(*Round your response to four decimal places*)

Is the estimated intercept $\hat{\beta}_0$ meaningful in this case?

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

Interpret the estimated slope $\hat{\beta}_1$.

- **A.** If a worker's age increases by 1 year, earnings increase, on average, by 99.07%.
- **B.** If a worker's age increases by 0.9907 year, earnings increase, on average, by 1 dollars per hour.
- **C.** If a worker's age increases by 0.9907 year, earnings increase, on average, by $0.9907 dollars per hour.
- **D.** If a worker's age increases by 1 year, earnings increase, on average, by $0.9907 dollars per hour.

Bob's is a 34-year-old worker. Predict Bob's earnings using the estimated regression.

Bob's predicted earnings are $[ ] dollars per hour.

(*Round your response to two decimal places*)

Alexis is a 30-year-old worker. Predict Alexis's earnings using the estimated regression.

Alexis's predicted earnings are $[ ] dollars per hour.

(*Round your response to two decimal places*)

Compute the R^2 for the regression above.

The R^2 for the regression above is [ ]

(*Round your response to four decimal places*)

Does age account for a large fraction of the variance in earnings across individuals?

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

Answers -11.6046

0.9907

B. No.

D. If a worker's age increases by 1 year, earnings increase, on average, by $0.9907 dollars per hour.

22.08

18.12

0.0878

B. No.

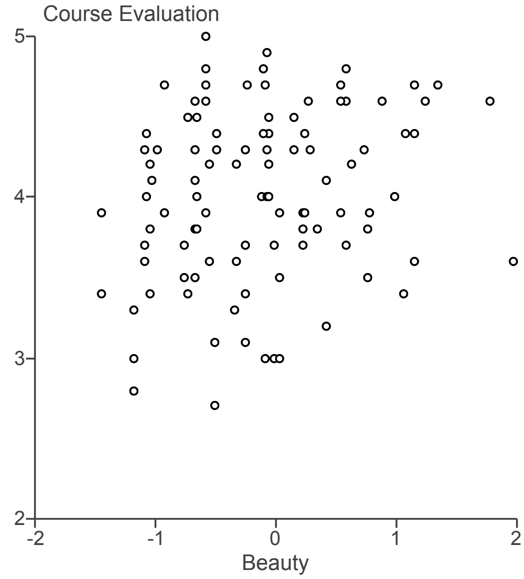
ID: General Empirical 4.1

---

25. One of the characteristics is an index of the professor's "beauty" as rated by a panel of six judges. In this exercise, you will investigate how course evaluations are related to the professor's beauty.

The following table uses data on course evaluations, course characteristics, and professor characteristics for 463 courses at the University of Texas at Austin. Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here [3]. Use a statistical package of your choice to answer the following questions.

Observe the following scatterplot of average course evaluations (*Course Evaluations*) on the professor's beauty (*Beauty*).



Does there appear to be a relationship between the variables?

- **A.** Yes, there appears to be a positive relationship between the variables.
- **B.** There is no evident relationship between the variables.
- **C.** Yes, there appears to be a negative relationship between the variables.
- **D.** There appears to be a positive relationship between the variables at low levels of professor's beauty, but a negative relationship at

Suppose you are interested in estimating the following model

$$\textit{Course Evaluation} = \beta_0 + \beta_1 \textit{Beauty} + u$$

Run a regression of average course evaluation (*Course Evaluation*) on the professor's beauty (*Beauty*).

What is the estimated intercept $\hat{\beta}_0$?

The estimated intercept $\hat{\beta}_0$ is ______.

(*Round your response to three decimal places*)

What is the estimated slope $\hat{\beta}_1$?

The estimated slope $\hat{\beta}_1$ is ______.

(*Round your response to three decimal places*)

Is the estimated intercept $\hat{\beta}_0$ meaningful in this case?

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

Professor Watson has an average value of *Beauty,* while Professor Stock's value of *Beauty* is one standard deviation above the average. Predict Professor Stock's and Professor Watson's course evaluations.

Professor Watson's predicted value of course evaluation is [ ].

(*Round your response to two decimal places*)

Professor Stock's predicted value of course evaluation is [ ].

(*Round your response to two decimal places*)

The standard deviation of course evaluation is 0.546, while the standard deviation of beauty is 0.743. Is the estimated effect of *Beauty* on *Course Evaluation* large or small?

- **A.** Large, because a one standard deviation increase in beauty is expected to increase course evaluation by 0.13.
- **B.** Small, because a one standard deviation increase in beauty is expected to increase course evaluation by 4.026.
- **C.** Small, because a one standard deviation increase in beauty is expected to increase course evaluation by 0.13.
- **D.** Large, because a one standard deviation increase in beauty is expected to increase course evaluation by 4.026.

Compute the R^2 for the regression above.

The R^2 for the regression above is [ ]

(*Round your response to four places*)

Does *Beauty* explain a large fraction of the variance in evaluations across courses?

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

3: More Info

Variable Definitions

Variable	Definition
CourseEval	"Course overall" teaching evaluation score, on a scale of 1 (very unsatisfactory) to 5 (excellent).
Beauty	Rating of instructor physical appearance by a panel of six students, averaged across the six panelists, shifted to have mean zero.

Answers A. Yes, there appears to be a positive relationship between the variables.

4.026

0.170

A. Yes.

4.00

4.13

C. Small, because a one standard deviation increase in beauty is expected to increase course evaluation by 0.13.

0.0536

B. No.

ID: General Empirical 4.2

26. Enter text here...Using the Excel data set, Teachers_rating.xls[4], described[5] in Empirical Exercise 4.2, create a scatter plot with *course evaluation* on the vertical axis and *beauty* on the horizontal axis. Right click any data point and click the option to "Add trend line." In the format option that pops up check the circle that will fit a linear trend and then check the boxes that will display the equation of the line and display the R^2. When that is done, answer the first three questions:

The equation of the trend line indicates that

- **A.** a one unit increase in the beauty index decreases the teacher's rating by -3.99 points
- **B.** a one unit increase in the beauty index increases the teacher's rating by 3.99 points
- **C.** a one unit increase in the beauty index increases the teacher's rating by 0.133 points
- **D.** a one unit increase in the beauty index decreases the teacher's rating by 0.133 points

A teacher with a score of 0.50 on the beauty index would receive what score on her teaching evaluation?

- **A.** 5.33
- **B.** -2.33
- **C.** 4.06
- **D.** -4.33

An R^2 of 0.0357 indicates that

- **A.** 3.57 percent of the variation in the teaching evaluation score can be explained by variation in the beauty index score
- **B.** 35.7 percent of the variation in the teaching evaluation score can be explained by variation in the beauty index score
- **C.** 35.7 percent of the variation in the beauty index can be explained by variation in the teaching evaluation score.
- **D.** 3.57 percent of the variation in the beauty index can be explained by variation in the teaching evaluation score

Use the Excel Regression module (Data/Data Analysis/Regression) and regress the course evaluation score on the beauty rating. What is the intercept in the fitted regression line and how do you interpret it? (hint: get the mean of the beauty score).

- **A.** 3.997 and it is the expected course evaluation score when the beauty score is set 1
- **B.** 3.997 and it is the expected course evaluation score when the beauty score is set 0
- **C.** 3.997 and it is the expected course evaluation score when the beauty score is set to its mean
- **D.** Both b and c are correct.

Assume the R^2 and the adjusted R^2 on the output were erased in an exam. How could you compute the R^2 (not the adjusted R^2) from the remaining elements in the regression output?

- **A.** Divide SSR by TSS
- **B.** Subtract (SSR/TSS) from 1?
- **C.** Divide the ESS by the TSS?
- **D.** Both (a) and (b)

Create a scatter plot as in part (1) with the *teaching evaluation* on the vertical axis. But this time put *age* (Professor's Age) on the horizontal axis. As in (1), add a linear trend line to the scatter and show the equation and R^2. Answer the following questions.

Based solely on a visual inspection of the trend line, you would initially conclude that

- **A.** There is a strong negative linear relationship between age and the teaching evaluation.
- **B.** None of the above
- **C.** There is no obvious linear association between age and the teaching evaluation.
- **D.** There is a strong positive linear relationship between age and the teaching evaluation.

Based on the R^2 between age and the teaching evaluation in the this plot you would conclude that

- A. Both (b) and (c) are correct.
- B. *Age* explains more of the variation in the teaching evaluation than does the beauty score in part (1)
- C. *Age* explains less of the variation in the teaching evaluation than does the beauty score in part (1)
- D. *Age* explains 29 percent of the variation in age where as the beauty score explain 37.5 percent of the variation in the teaching evalu

Using the Excel regression module, regress the teaching evaluation on the professor's age. Based on the output,

- A. The standard error of the estimate is an estimator of the standard deviation of the error term.
- B. The standard error of the estimate is the square root of the SSR divided by n-2
- C. The standard error of the estimate is 0.555
- D. All the above are correct

Consider the regression output in part (8). Let Y_i be the teaching evaluation for professor "i", and let X_i be the professor "i" age. Let $\hat{\beta}_0$ ar $\hat{\beta}_1$ be the estimated coefficients. Thus, the following sum, $\sum_i \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$ is equal to

- A. 1.235
- B. 1.235
- C. 141.86
- D. 142.23

Based on the regression output in part (8), the difference in the teaching evaluation of a professor 40 years of age versus a professor 30 ye of age would be

- A. 0.00293
- B. -0.00293
- C. -0.0293
- D. 0.2930

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

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

Answers C. a one unit increase in the beauty index increases the teacher's rating by 0.133 points

C. 4.06

A. 3.57 percent of the variation in the teaching evaluation score can be explained by variation in the beauty index score

D. Both b and c are correct.

D. Both (a) and (b)

C. There is no obvious linear association between age and the teaching evaluation.

C. *Age* explains less of the variation in the teaching evaluation than does the beauty score in part (1)

D. All the above are correct

C. 141.86

C. -0.0293

ID: General Empirical 4.2 (static)

27. In this exercise, you will use these data to investigate the relationship between the number of completed years of education for young adults and the distance from each student's high school to the nearest four-year college. (Proximity lowers the cost of education, so that students who live closer to a four-year college should, on average, complete more years of higher education.)

The following table contains data from a random sample of high school seniors interviewed in 1980 and re-interviewed in 1986. Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the dataset is available here [6]. Use a statistical package of your choice to answer the following questions.

Suppose you are interested in estimating the following model

$$ED = \beta_0 + \beta_1 Dist + u$$

Run a regression of years of completed education (*ED*) on distance to the nearest college (*Dist*), where *Dist* is measured in tens of miles. (For example, *Dist* = 2 means that the distance is 20 miles).

What is the estimated intercept $\hat{\beta}_0$?

The estimated intercept $\hat{\beta}_0$ is ______.

(*Round your response to three decimal places*)

What is the estimated slope $\hat{\beta}_1$?

The estimated slope $\hat{\beta}_1$ is ______.

(*Round your response to three decimal places*)

Is the estimated intercept $\hat{\beta}_0$ meaningful in this case?

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

How does the average value of years of completed schooling change when colleges are built close to where students go to high school?

- **A.** The regression predicts that if colleges are built 10 miles closer to where students go to high school, average years of college will
- **B.** The regression predicts that if colleges are built 10 miles closer to where students go to high school, average years of college will
- **C.** The regression predicts that if colleges are built 10 miles closer to where students go to high school, average years of college will
- **D.** The regression predicts that if colleges are built 10 miles closer to where students go to high school, average years of college will

Bob's high school was 43 miles from the nearest college. Predict Bob's years of completed education using the estimated regression.

Bob's predicted years of education completed is ______.

(*Round your response to two decimal places*)

John's high school was 50 miles from the nearest college. Predict John's years of completed education using the estimated regression.

John's predicted years of education completed is ______.

(*Round your response to two decimal places*)

Compute the R^2 for the regression above.

The R^2 for the regression above is ______

(*Round your response to four decimal places*)

Does distance to college explain a large fraction of the variance in educational attainment across individuals?

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

Compute the value of the standard error of the regression and specify its units.

The standard error of the regression (*SER*) is [ ] (1) ________.

(*Round your response to four decimal places*)

6: More Info

Variable Definitions

Variable	Definition
Dist	Distance from 4yr College in 10's of miles.
Ed	Years of Education Completed. Rouse computed years of education by assigning 12 years to all members of the senior class. Each additional year of secondary education counted as a one year. Students with vocational degrees were assigned 13 years, AA degrees were assigned 14 years, BA degrees were assigned 16 years, those with some graduate education were assigned 17 years, and those with a graduate degree were assigned 18 years.

(1) ○ years
○ tens of miles
○ months

Answers 13.945

−0.003

A. Yes.

A.
The regression predicts that if colleges are built 10 miles closer to where students go to high school, average years of college wi increase by 0.003 years.

13.93

13.93

0.0000

B. No.

3.6290

(1) years

ID: General Empirical 4.3

---

28. The regression R^2 is a measure of:

- ○ **A.** the square of the determinant of *R*.
- ○ **B.** whether or not *ESS* > *TSS*.
- ○ **C.** whether or not *X* causes *Y*.
- ○ **D.** the goodness of fit of your regression line.

Answer: D. the goodness of fit of your regression line.

ID: Test A Ex 4.3.4

---

29. In a sport like basketball, there is always a preference given to taller players since it is widely believed that taller players tend to perform better than shorter players. However, if one looks at the history of NBA, there have been instances when players with relatively shorter heights have performed exceedingly well. Michael, an ardent follower of this game, is interested in understanding the relationship between a player's height and how well he performs. Accordingly, he collects data on the heights of 105 shooting guards and the number of points they scored in a tournament. The estimated regression function is:

$$\hat{P}_i = 1.20 + 0.91H_i,$$

where $\hat{P}_i$ is the predicted value of the number points scored by the ith player in the tournament and H_i is the height of the ith player measured in inches.

Michael wants to check how well his OLS regression line fits the data before arriving at any conclusion. From the sample data, he makes the following calculations.

$$\sum_{i=1}^{105} \left(\hat{P}_i - \overline{P}\right)^2 = 80.45.$$

$$\sum_{i=1}^{105} \left(P_i - \overline{P}\right)^2 = 103.47.$$

The regression R^2 is ☐.

(Round your answer to two decimal places.)

Which of the following statements are true about the value of R^2? (*Check all that apply*).

☐ **A.** The value of R^2 lies between 0 and 1.

☐ **B.** If the independent variable explains none of the variations in the dependent variable, then the value of R^2 will be 0.

☐ **C.** If the independent variable explains all the variations in the dependent variable, then the value of R^2 will be -1.

☐ **D.** The value of R^2 lies between -1 and $+1$.

Answers 0.78

A. The value of R^2 lies between 0 and 1., B.
If the independent variable explains none of the variations in the dependent variable, then the value of R^2 will be 0.

ID: Concept Exercise 4.3.1

30. A researcher estimates a regression to test the effect of educational attainment (X_i) on the average earnings of individuals (Y_i).

The sample correlation coefficient (r) between years of educational attainment and the average earnings of individuals from a sample of 85 individuals was calculated to be 0.98.

The value of the regression R^2 will be ______.

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

The value of R^2 indicates that the regressor (years of educational attainment) is (1) ______ at predicting the regressand (average earnings of individuals).

(1) ○ good
○ not very good

Answers 0.96

(1) good

ID: Concept Exercise 4.3.2

---

31. A salon is currently located far-off from the main city. The owner is considering to shift his salon close to the main city as it would increase his revenues. However, he supposes that as he moves closer to the city, he would have to pay a higher rent. He collects data by surveying 150 prospective locations and estimates the OLS regression:

$$\widehat{PR}_i = 2.3 - 1.2d_i,$$

where $\widehat{PR}_i$ represents the predicted extra rent (measured in thousands of dollars) he has to pay corresponding to its distance (d_i) (measured in kilometers or km) between the salon and the main city. From the sample data, he makes the following calculations.

$$\sum_{i=1}^{150} \left(PR_i - \overline{PR}\right)^2 = 170.54.$$

$$\sum_{i=1}^{150} \left(PR_i - \widehat{PR}\right)^2 = 27.25.$$

According to the estimated regression function, if the owner moves one km towards the city he has to pay an extra rent of $1,200. He wants to check how well the regression line fits the data before deciding to move his salon to the city.

The regression R^2 is ______.

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

If the value of the estimated slope coefficient in the regression equation was found to be zero, then what would have been the value of R^2?

○ **A.** 0.

○ **B.** 0.5.

○ **C.** 1.

○ **D.** Insufficient information provided.

Answers 0.84

A. 0.

ID: Concept Exercise 4.3.3

---

32. A group of randomly selected 85 automobile dealers across the automobile industry was surveyed and information about their advertisement expenditure and the total number of cars they sold in a particular year was collected. The estimated OLS regression is:

$$\widehat{Q}_i = 3.4 + 0.09E_i,$$

where $\widehat{Q}_i$ and E_i denote the quantity of automobiles sold by the ith dealer and their advertisement expenditure, respectively.

Calculations show that the regression $R^2 = 0.15$ and

$$\sum_{i=1}^{85} \left(Q_i - \overline{Q}\right)^2 = 436.45.$$

Based on the given information, the standard error of regression is ______.

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

Which of the following statements hold true about the standard error of regression?

- ☐ **A.** The standard error of the regression is an estimator of the standard deviation of the regression error.
- ☐ **B.** The standard error of the regression is an estimator of the standard error of the dependent variable.
- ☐ **C.** The standard error of the regression and the dependent variable are measured in the same unit.
- ☐ **D.** The standard error of the regression and the regressor have the same measure of unit.

Answers 2.11

A. The standard error of the regression is an estimator of the standard deviation of the regression error., C. The standard error of the regression and the dependent variable are measured in the same unit.

ID: Concept Exercise 4.3.4

33. Suppose you are interested in studying the relationship between education and wage. More specifically, suppose that you believe the relationship to be captured by the following linear regression model,

$$Wage = \beta_0 + \beta_1 Education + u$$

Suppose further that the only unobservable that can possibly affect both wage and education is intelligence of the individual.

OLS assumption (1): The conditional distribution of u_i given X_i has a mean of zero. Mathematically, $E\left(u_i|X_i\right) = 0$.

Which of the following provides evidence in favor of OLS assumption #1? (*Check all that apply*)

- ☐ **A.** $covariance(Intelligence, Education) \neq 0$.
- ☐ **B.** $corr(Intelligence, Education) \neq 0$.
- ☐ **C.** $E(Intelligence|Education = x) = E(Intelligence|Education = y)$ for all $x \neq y$.
- ☐ **D.** $corr(Intelligence, Education) = 0$.

Which of the following provides evidence against of OLS assumption #1? (*Check all that apply*)

- ☐ **A.** $corr(Intelligence, Education) \neq 0$.
- ☐ **B.** $E(Intelligence|Education = x) = E(Intelligence|Education = y)$ for all $x \neq y$.
- ☐ **C.** $covariance(Intelligence, Education) \neq 0$.
- ☐ **D.** $corr(Intelligence, Education) = 0$.

OLS assumption (2): $\left(X_i, Y_i\right)$, $i = 1,\ldots, n$ are independently and identically distributed.

Suppose you would like to draw a sample to study the effect of education on wage. Which of the following provides evidence in favor of OLS assumption #2? (*Check all that apply*)

- ☐ **A.** A sample consisting of a group of college students is drawn repeatedly each year over the course of their college careers.
- ☐ **B.** $corr(Intelligence, Education) = 0$.
- ☐ **C.** A random sample is drawn from a population of college graduates.
- ☐ **D.** A sample consisting of all honor students is drawn from a population of college graduates.

Suppose you would like to draw a sample to study the effect of education on wage. Which of the following provides evidence against OLS assumption #2? (*Check all that apply*)

- ☐ **A.** Observations consisting of the same group of college students are drawn repeatedly each year over the course of their college ca
- ☐ **B.** A sample consisting of all honor students is drawn from a population of college graduates.
- ☐ **C.** A random sample is drawn from a population of college graduates.
- ☐ **D.** $corr(Intelligence, Education) = 0$.

OLS assumption (3): Large outliers are unlikely. Mathematically, X and Y have nonzero finite fourth moments: $0 < E\left(X_i^4\right) < \infty$ and $0 < E\left(Y_i^4\right) < \infty$.

Suppose you would like to draw a sample to study the effect of education on wage. Which of the following provides evidence in favor OLS assumption #3? (*Check all that apply*)

- ☐ **A.** The maximum wage an individual can get is a finite number.
- ☐ **B.** The years of education an individual can get is bounded above.
- ☐ **C.** Half of the wages in the sample were incorrectly multiplied by 1 million when recorded.
- ☐ **D.** For some individuals in the sample, years of education were recorded in days rather than years.

Suppose you would like to draw a sample to study the effect of education on wage. Which of the following provides evidence against OLS assumption #3? (*Check all that apply*)

- A. For some individuals in the sample, years of education were recorded in days rather than years.
- B. The maximum wage an individual can get is a finite number.
- C. Half of the wages in the sample were incorrectly multiplied by 1 million when recorded.
- D. The years of education an individual can get is bounded above.

Answers C. $E(Intelligence|Education = x) = E(Intelligence|Education = y)$ for all $x \neq y$.

A. $corr(Intelligence, Education) \neq 0$., C. $covariance(Intelligence, Education) \neq 0$.

C. A random sample is drawn from a population of college graduates.

A.
Observations consisting of the same group of college students are drawn repeatedly each year over the course of their college careers
, B. A sample consisting of all honor students is drawn from a population of college graduates.

A. The maximum wage an individual can get is a finite number., B.
The years of education an individual can get is bounded above.

A. For some individuals in the sample, years of education were recorded in days rather than years. , C.
Half of the wages in the sample were incorrectly multiplied by 1 million when recorded.

ID: Review Concept 4.2

34. A professor decides to run an experiment to measure the effect of time pressure on final exam scores. He gives each of the 400 students in his course the same final exam, but some students have 90 minutes to complete the exam while others have 120 minutes. Each student is randomly assigned one of the examination times based on the flip of a coin. Let Y_i denote the number of points scored on the exam by the i^{th} student ($0 \le Y_i \le 100$), let X_i denote the amount of time that the student has to complete the exam ($X_i = 90$ or 120), and consider the regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ E(u_i) = 0$$

Which of the following are true about the unobservable u_i? (*Check all that apply*)

☐ **A.** u_i will be zero for all students because time spent studying is likely the only factor that affects exam performance.

☐ **B.** u_i represents factors other than time that influence the student's performance on the exam.

☐ **C.** Different students will have different values of u_i because they have unobserved individual specific traits that affect exam performa

☐ **D.** All students will necessarily have the same value of u_i because they are part of the same population.

The Least Squares Assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, \ldots, n \text{ where}$$

1. The error term u_i has conditional mean zero given X_i: $E(u_i|X_i) = 0$;
2. (X_i, Y_i), $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) draws from their joint distribution; and
3. Large outliers are unlikely: X_i and Y_i have nonzero finite fourth moments.

Assuming this year's class is a typical representation of the same class in other years, are OLS assumption (2) and (3) satisfied?

○ **A.** Both OLS assumption #2 and OLS assumption #3 are satisfied.

○ **B.** Neither OLS assumption #2 nor OLS assumption #3 is satisfied.

○ **C.** Only OLS assumption #3 is satisfied.

○ **D.** Only OLS assumption #2 is satisfied.

The estimated regression is

$$\widehat{Y}_i = 34 + 0.84 X_i$$

Compute the estimated regression's prediction for the average score of students given 97, 123, or 159 minutes to complete the exam.

Given 97 minutes, the estimated regression's prediction for the average score of students is ______.

Given 123 minutes, the estimated regression's prediction for the average score of students is ______.

Given 159 minutes, the estimated regression's prediction for the average score of students is ______.

(Round your responses to two decimal places.)

Compute the estimated gain in score for a student who is given an additional 47 minutes on the exam.

The estimated gain in score for a student who is given an additional 47 minutes on the exam is ______

(Round your response to two decimal places.)

Answers B. u_i represents factors other than time that influence the student's performance on the exam., C. Different students will have different values of u_i because they have unobserved individual specific traits that affect exam performance.

A. Both OLS assumption #2 and OLS assumption #3 are satisfied.

115.48

137.32

167.56

39.48

ID: Exercise 4.5

---

35. Show that the first least squares assumption, $E\left(u_i|X_i\right) = 0$, implies that $E\left(Y_i|X_i\right) = \beta_0 + \beta_1 X_i$.

Using $E\left(u_i|X_i\right) = 0$, we have

$$E\left(Y_i|X_i\right) = E\left(\beta_0 + \beta_1 X_i + u_i|X_i\right) = (1)\_\_\_\_\_\_ + (2)\_\_\_\_\_\_ E(X_i|X_i) + E((3)\_\_\_\_\_\_|X_i)$$

(1) ○ β_0 ○ β_0^2 ○ 0

(2) ○ β_1 ○ β_1^2 ○ 0

(3) ○ u_i ○ u_i^2 ○ $\frac{u_i^2}{2}$

Answers (1) β_0

(2) β_1

(3) u_i

ID: Exercise 4.6

---

36. **The Least Squares Assumptions**

$$Y_i = \beta_0 + \beta_0 X_i + u_i,\ i = 1, \ldots, n, \text{ where}$$

1. The error term u_i has conditional mean zero given X_i: $E\left(u_i|X_i\right) = 0$;
2. $\left(X_i, Y_i\right)$, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) draws from their joint distribution; and
3. Large outliers are unlikely: X_i and Y_i have nonzero finite fourth moments.

Suppose the first assumption is replaced with $E\left(u_i|X_i\right) = 2$. What happens to $E\left(Y_i|X_i\right)$?

- **A.** Nothing changes.
- **B.** Both the intercept β_0 and the slope β_1 change to $\beta_0 + 2$ and $\beta_1 + 2$ respectively.
- **C.** The intercept β_0 changes to $\beta_0 + 2$.
- **D.** The slope β_1 changes to $\beta_1 + 2$.

Are the rest of the OLS assumptions satisfied?

- **A.** Neither OLS assumption (2) nor (3) is satisfied.
- **B.** OLS assumption (3) is satisfied but not (2).
- **C.** OLS assumption (2) is satisfied but not (3).
- **D.** Both OLS assumptions (2) and (3) are satisfied.

Answers C. The intercept β_0 changes to $\beta_0 + 2$.

D. Both OLS assumptions (2) and (3) are satisfied.

ID: Exercise 4.8

---

37. $E\left(\mu_i \,|X_i\right) = 0$ says that:

- **A.** dividing the error by the explanatory variable results in a zero (on average).
- **B.** the sample mean of the Xs is much larger than the sample mean of the errors.
- **C.** the sample regression function residuals are unrelated to the explanatory variable.
- **D.** the conditional distribution of the error given the explanatory variable has a zero mean.

Answer: D. the conditional distribution of the error given the explanatory variable has a zero mean.

ID: Test A Ex 4.4.5

---

38. Suppose you want to estimate the effect that number of years of schooling have on a person's earnings. Let *X* be the number of years of schooling a person has received and *Y* be this person's monthly income. The model you want to test would be:

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

where u_i is the error term, which could include factors like the ability of the person, ranking of the schools the person attended, income levels of the person's parents, etc. Suppose you decide to use a computerized random number generator that uses no information about the subject to select the individuals to include in your sample.

Which of the following statements are true? (*Check all that apply*.)

- ☐ **A.** If *X* and *Y* have kurtosis values of 5 and 7 respectively, large outliers in your data are unlikely.
- ☐ **B.** The conditional mean of u_i given X_i will be zero in this case.
- ☐ **C.** This sampling scheme produces i.i.d. observations on (X_i, Y_i).
- ☐ **D.** If X_i and u_i are uncorrelated, the conditional mean of u_i given X_i will be zero.

Suppose you decide to collect data for the same set of individuals over a period of 30 years to estimate the relationship between *X* and *Y*.

Which of the following statements is true in this case?

- ○ **A.** The conditional mean of u_i given X_i will continue to be zero as in the previous case.
- ○ **B.** The sampling scheme will no longer produce i.i.d. observations on (X_i, Y_i) as opposed to the previous case.
- ○ **C.** The conditional mean of u_i given X_i will no longer be zero as opposed to the previous case.
- ○ **D.** The sampling scheme will continue to produce i.i.d. observations on (X_i, Y_i) as in the previous case.

Answers A. If *X* and *Y* have kurtosis values of 5 and 7 respectively, large outliers in your data are unlikely., C. This sampling scheme produces i.i.d. observations on (X_i, Y_i).

B. The sampling scheme will no longer produce i.i.d. observations on (X_i, Y_i) as opposed to the previous case.

ID: Concept Exercise 4.4.1

---

39. Which of the following statements hold true for the sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ under the least squares assumptions? (*Check all that apply*.)

☐ **A.** $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively.

☐ **B.** The sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ is well approximated by the bivariate normal distribution if the sample is sufficiently large.

☐ **C.** $\hat{\beta}_0$ and $\hat{\beta}_1$ are biased estimators of β_0 and β_1, respectively.

☐ **D.** The sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ is always well approximated by the bivariate normal distribution.

Suppose a researcher conducts an experiment with a sample size of $n = 50$.

In this case, the normal approximations to the distributions of the OLS estimators of the researcher's regression parameters

(1) ____________ be reliable.

(1) ○ will not
○ will
○ may not

Answers A. $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively., B. The sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ is well approximated by the bivariate normal distribution if the sample is sufficiently large.

(1) may not

ID: Concept Exercise 4.5.1

40. Which of the following are the roles of the least square assumptions? (*Check all that apply*.)

- ☐ **A.** The least square assumptions help estimate the values of the population parameters with certainty.
- ☐ **B.** If the least square assumptions hold, then mathematically, the sampling distributions of OLS estimators are normal in large samples.
- ☐ **C.** If the least square assumptions hold, then X_i will explain all of the variation of Y_i in large samples.
- ☐ **D.** The least square assumptions organize the circumstances that pose difficulties for OLS regression.

In a large-sample where $\hat{\beta}_1$ is distributed normally as $N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$, which of the following statements hold true for the variance $\sigma^2_{\hat{\beta}_1}$?

(*Check all that apply*.)

- ☐ **A.** As the variance of X gets larger, the variance of $\hat{\beta}_1$ gets larger.
- ☐ **B.** As the variance of the error term gets smaller, the variance of $\hat{\beta}_1$ gets larger.
- ☐ **C.** As the variance of X gets larger, the variance of $\hat{\beta}_1$ gets smaller.
- ☐ **D.** As the variance of the error term gets smaller, the variance of $\hat{\beta}_1$ gets smaller.

Answers B.

If the least square assumptions hold, then mathematically, the sampling distributions of OLS estimators are normal in large samples.

, D. The least square assumptions organize the circumstances that pose difficulties for OLS regression.

C. As the variance of X gets larger, the variance of $\hat{\beta}_1$ gets smaller., D.

As the variance of the error term gets smaller, the variance of $\hat{\beta}_1$ gets smaller.

ID: Concept Exercise 4.5.2