Student: ____________________ **Date:** ____________________	**Instructor:** Richeng Piao **Course:** ECON 2560 - Applied Econometrics	**Assignment:** Practice Problem Set 5

1. Suppose you are interested in investigating the wage gender gap using data on earnings of men and women. Which of the following models best serves this purpose?

   - **A.** $Female = \beta_0 + \beta_1 Wage + u$ where *Female* (=1 if female) is an indicator variable and *u* the error term.
   - **B.** $Wage = \beta_0 + \beta_1 Female + u$, where *Female* (=1 if female) is an indicator variable and *u* the error term.
   - **C.** $Wage = \beta_0 + u$ where *u* is the error term.
   - **D.** $Male = \beta_0 + \beta_1 Female + u$ where *Male* (=1 if male) is an indicator variable and *u* the error term.

   Consider the regression model

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

   Where *Female* (=1 if female) is an indicator variable and *u* the error term.

   Identify the dependent and independent variables in the regression model above.

   Wage is the (1) __________ variable, while *Female* is the (2) __________ variable.

   (1) ○ dependent ○ independent
   (2) ○ dependent ○ independent

   Answers B. $Wage = \beta_0 + \beta_1 Female + u$, where *Female* (=1 if female) is an indicator variable and *u* the error term.

   (1) dependent

   (2) independent

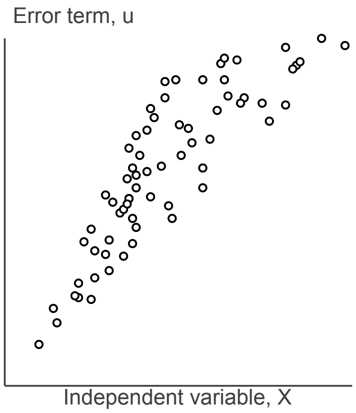
   ID: Review Concept 5.2

2. Consider the regression model below and let (X_i, Y_i), $i = 1,..., n$ be an i.i.d. set of observations.

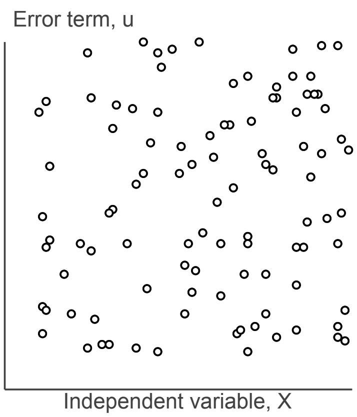
$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

The error u_i is *homoskedastic* if the variance of the conditional distribution of u_i given X_i is constant for $i = 1,..., n$ and in particular does not depend on X_i. Otherwise, the error term is *heteroskedastic*.

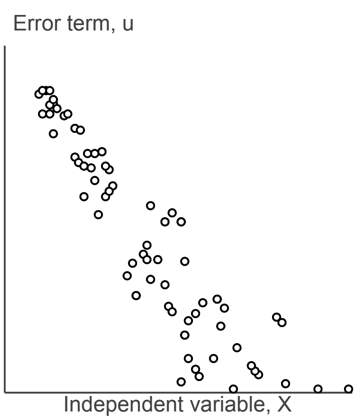
Determine whether the following scatter plots depict homoskedastic or heteroskedastic errors u_i.

1. 

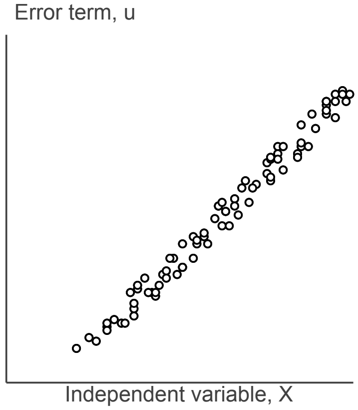
(1) ______________.

2. 

(2) ______________.

3. 

(3) ______________

4. 

(4) ____________.

(1) ○ homoskedastic ○ heteroskedastic
(2) ○ homoskedastic ○ heteroskedastic
(3) ○ homoskedastic ○ heteroskedastic
(4) ○ homoskedastic ○ heteroskedastic

Answers (1) heteroskedastic

(2) homoskedastic

(3) heteroskedastic

(4) homoskedastic

ID: Review Concept 5.3

3. Suppose that a researcher, using data on class size (*CS*) and average test scores from 100 third-grade classes, estimates the OLS regression

$$\widehat{TestScore} = \underset{(19.3800)}{494.3800} + \underset{(2.3647)}{(-5.5290)} \times CS, \quad R^2 = 0.07, \quad SER = 10.9$$

Construct a 95% confidence interval for β_1, the regression slope coefficient.

The 95% confidence interval for β_1, the regression slope coefficient, is (☐ , ☐). (*Round your responses to two decimal places.*)

The *t*-statistic for the two-sided test of the null hypothesis H_0: $\beta_1 = 0$ is ☐. (*Round your response to four decimal places.*)

Note: Assume a normal distribution.

The *p*-value for the two-sided test of the null hypothesis H_0: $\beta_1 = 0$ is ☐. (*Round your response to four decimal places.*)

Do you reject the null hypothesis at the 1% level?

- **A.** No, because the *p*-value is greater than 0.01.
- **B.** Yes, because the *p*-value is less than 0.01.
- **C.** Yes, because the *t*-statistic is greater than 2.58.
- **D.** Yes, because the *t*-statistic is less than 2.58.

The *p*-value for the two-sided test of the null hypothesis H_0: $\beta_1 = -5.3$ is ☐. (*Round your response to four decimal places.*)

Without doing any additional calculations, determine whether -5.3 is contained in the 95% confidence interval for β_1.

- **A.** Yes, -5.3 is contained in the 95% confidence interval for β_1.
- **B.** No, -5.3 is not contained in the 95% confidence interval for β_1.

The 99% confidence interval for β_0 is (☐ , ☐). (*Round your responses to one decimal place.*)

Answers -10.16

-0.89

-2.3381

0.0194

A. No, because the *p*-value is greater than 0.01.

0.9228

A. Yes, -5.3 is contained in the 95% confidence interval for β_1.

444.4

544.4

ID: Exercise 5.1

4. The *t*-statistic is calculated by dividing:

- **A.** the estimator minus its hypothesized value by the standard error of the estimator.
- **B.** the OLS estimator by its standard error.
- **C.** the slope by 1.96.
- **D.** the slope by the standard deviation of the explanatory variable.

Answer: A. the estimator minus its hypothesized value by the standard error of the estimator.

ID: Test A Ex 5.1.1

---

5. Consider the estimated equation from your textbook:

$$\widehat{Test\ Score} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28} \times STR,\ R^2 = 0.051,\ SER = 18.6$$

The *t*-statistic for the slope is approximately:

- **A.** 1.76.
- **B.** 0.52.
- **C.** 67.20.
- **D.** 4.38.

Answer: D. 4.38.

ID: Test A Ex 5.1.2

---

6. Imagine that you were told that the *t*-statistic for the slope coefficient of the regression line $\widehat{Test\ Score} = 698.9 - 2.28 \times STR$ was 4.38.

What are the units of measurement for the *t*-statistic?

- **A.** $\frac{Test\ Score}{STR}$.
- **B.** Number of students per teacher.
- **C.** Points of the test score.
- **D.** Standard deviations.

Answer: D. Standard deviations.

ID: Test B Ex 5.1.1

---

7. A professor wants to investigate the relationship between the grades students obtain in their midterm exam (X_i) and the grades they obtain (Y_i) in the final exam.

The professor collects data from 65 randomly chosen students. The estimated OLS regression is:

$$\hat{Y}_i = 52 + 0.85X_i,$$

where $\hat{Y}_i$ denotes the predicted value of the grades obtained in the final exam by the i^{th} individual and X_i denotes the grades obtained in the midterm exam.

From the sample data he makes the following calculations:

$$\sum_{i=1}^{65} \left(X_i - \bar{X}\right)^2 = 250.12,$$

$$\sum_{i=1}^{65} \left(X_i - \bar{X}\right)^2 \hat{u}_i^2 = 410.25,$$

where $\hat{u}_i^2$ is the square of the residual for the i^{th} observation.

He wants to test whether the grades obtained in the midterm exam have any effect on the grades obtained in the final exam or not.

Which of the following are the null and the alternative hypotheses of the test the professor wishes to conduct?

- **A.** $H_0: \beta_1 = 0.85$ vs. $H_1: \beta_1 \neq 0.85$.
- **B.** $H_0: \beta_1 \neq 0$ vs. $H_1: \beta_1 = 0$.
- **C.** $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.
- **D.** $H_0: \beta_1 \neq 0.85$ vs. $H_1: \beta_1 = 0.85$.

If $\hat{\beta}_1$ is the estimated slope coefficient, then the value of standard error of the estimated slope ($SE(\hat{\beta}_1)$) is ______.

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

Answers C. $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.

0.0825

ID: Concept Exercise 5.1.1

8. A study conducted by a group of yoga professionals claim that introducing meditation activities in workplace can significantly increase the performance of the employees. After finding out about this study, the HR department of company ABC introduces similar activities in their office and encourages every employee to participate. At the end of one month the department is interested in checking the claims of the study. The department selects a random sample of 100 employees and collects information on the number of hours they have spent meditating (X) and they score their performances based on their revenue contribution to the company (Y_i). They estimate the following regression function:

$$\hat{Y} = 45.63 + 3.60X,$$

where $\hat{Y}$ denotes the predicted value of the score obtained by the i^{th} individual and X denotes the number of hours they spend meditating. The $SE(\hat{\beta}_1)$ is 4.54.

The HR department wishes to test whether or not meditating increases the performance of their employees.

Which of the following are the null and the alternative hypotheses of the test the department wishes to conduct?

- **A.** $H_0: \beta_1 = 3.60$ vs. $H_1: \beta_1 \neq 3.60$.
- **B.** $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.
- **C.** $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 > 0$.
- **D.** $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 < 0$.

The value of the t-statistic associated with the test the HR department wishes to conduct is ______.

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

At the 5% significance level, the department would (1) ______ the null hypothesis.

(1) ○ reject
○ fail to reject

Answers C. $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 > 0$.

0.79

(1) fail to reject

ID: Concept Exercise 5.1.2

9. A researcher wants to test the relationship between the number of years of formal education received (X_i) and the average weekly earnings (Y_i) (measured in hundred dollars).

A report released by a government agency suggests that the average weekly earnings of individuals with no formal education is equal to $545. The researcher wants to test whether the average weekly earnings with no formal education is $545 or greater than that. He collects data from a sample of 150 individuals and estimates the following regression function:

$$\hat{Y}_i = 6.91 + 1.21X_i,$$
$$(3.24)\ (4.25)$$

where $\hat{Y}_i$ is the predicted value of the weekly earnings for the i^{th} individual and the standard errors for the coefficients appear in parenthesis.

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

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

At the 5% significance level, the researcher will (1) ______ the null hypothesis that the average weekly earnings of individuals with no formal education is equal to $545.

(1) ○ reject
○ fail to reject

Answers 0.45

(1) fail to reject

ID: Concept Exercise 5.1.3

---

10. A fertilizer manufacturing company has claimed that each extra unit of its fertilizer would increase the fruit bearing capacity of an apricot tree by 2.68 apricots. Wendy, an apricot farmer, wishes to test whether this increase in fruit bearing capacity is actually 2.68 apricots or not. She selects 100 trees at random from her orchard and uses this fertilizer on those trees and estimates the following regression:

$$\hat{Y}_i = 550 + 4.93X_i,$$

where $\hat{Y}_i$ denotes the predicted number of apricots obtained from the i^{th} tree and X_i denotes the number of units of fertilizer used on the i^{th} tree.

If β_0 and β_1 denote the intercept coefficient and the slope coefficient, respectively, which of the following options states Wendy's null and alternative hypotheses?

- **A.** $H_0: \beta_1 \geq 2.68$ and $H_1: \beta_1 < 2.68$.
- **B.** $H_0: \beta_0 = 4.93$ and $H_1: \beta_0 \neq 4.93$.
- **C.** $H_0: \beta_1 > 4.93$ and $H_1: \beta_1 \leq 4.93$.
- **D.** $H_0: \beta_1 = 2.68$ and $H_1: \beta_1 \neq 2.68$.

Suppose the standard error of the estimated slope is 0.89.

The *t*-statistic associated with the test Wendy wishes to conduct is ______.

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

The *t*-statistic suggests that at the 5% significance level, we (1) ______ the null hypothesis.

(1) reject
fail to reject

Answers D. $H_0: \beta_1 = 2.68$ and $H_1: \beta_1 \neq 2.68$.

2.53

(1) reject

ID: Concept Exercise 5.1.4

11. The chief of the emergency response services claims that their team responds to emergency distress calls within a 5 km radius in 360 seconds. An independent investigator wishes to test whether or not the presence of a thousand extra cars within the 5 km radius affects the emergency response time by 8.65 seconds. He randomly selects 100 distress calls and estimates the following regression:

$$\hat{Y}_i = 370 + 0.23X_i,$$

where $\hat{Y}_i$ denotes the predicted response time measured in seconds for the i^{th} emergency call and X_i denotes the number of cars (in thousands) on the road within a 5 km radius at the time of the i^{th} emergency call.

If β_1 is the slope coefficient, which of the following options states the investigator's null and alternative hypotheses?

- **A.** $H_0: \beta_1 > 0.23$ and $H_1: \beta_1 \le 0.23$.
- **B.** $H_0: \beta_1 = 8.65$ and $H_1: \beta_1 \ne 8.65$.
- **C.** $H_0: \beta_1 = 0.23$ and $H_1: \beta_1 \ne 0.23$.
- **D.** $H_0: \beta_1 > 8.65$ and $H_1: \beta_1 \le 8.65$.

Suppose $|t^{act}|$ associated with the test the investigator wishes to conduct is 1.98.

The *p*-value will be ______.

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

If the pre-specified significance level of the test was 5%, the calculated *p*-value suggests that we (1) ______ the null hypothesis.

(1) fail to reject
reject

Answers B. $H_0: \beta_1 = 8.65$ and $H_1: \beta_1 \ne 8.65$.

0.048

(1) reject

ID: Concept Exercise 5.1.5

12. Consider the regression model below and let (X_i, Y_i), $i = 1, \ldots, n$ be an i.i.d. set of observations.

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

Specify the order of the 3 steps required to test the null hypothesis that the slope coefficient β_1 equals zero? That is,

$$H_0 : \beta_1 = 0$$

A. Compute the standard error of the estimated slope coefficient $\widehat{\beta}_1^{act}$, $SE\left(\widehat{\beta}_1^{act}\right) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$. [ ]

B. Compute the $t-$stastistic, $t = \dfrac{\widehat{\beta}_1^{act} - 0}{SE\left(\widehat{\beta}_1^{act}\right)}$. [ ]

C. Compute the $p-$value, $\Pr_{H_0}\left(|t| > |t^{act}|\right)$. [ ]

(*Enter 1, 2 or 3 in each box*)

Answers 1

2

3

ID: Review Concept 5.1

---

13. Suppose that a random sample of 208 twenty-year-old men is selected from a population and that their heights and weights are recorded. A regression of weight on height yields

$$\widehat{Weight} = \underset{(2.2360)}{(-103.3864)} + \underset{(0.3224)}{4.0976} \times Height,\ R^2 = 0.842,\ SER = 10.6080$$

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

A man has a late growth spurt and grows 1.5600 inches over the course of a year. Construct a confidence interval of 99% for the person's weight gain.

The 99% confidence interval for the person's weight gain is ([ ], [ ]) (in pounds). (*Round your responses to two decimal places.*)

Answers 5.09

7.69

ID: Exercise 5.3

---

14. Suppose that (Y_i, X_i) satisfy the assumptions specified here. A random sample of n = 257 is drawn and yields

$$\underset{(1.8)}{\hat{Y} = 6.42} + \underset{(1.7)}{5.63X}, \quad R^2 = 0.21, \quad SER = 6.4$$

Where the numbers in parentheses are the standard errors of the estimated coefficients $\hat{\beta}_0 = 6.42$ and $\hat{\beta}_1 = 5.63$ respectively.

Suppose you wanted to test that β_1 is zero at the 5% level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the t-statistic and p-value for this test.

The t-statistic is ______.

(*Round your response to two decimal places*)

The p-value is ______.

(*Round your response to two decimal places*)

Based on the p-value computed above, would you reject the null hypothesis at the 5% level?

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

Construct a 95% confidencen interval for β_1.

The 95% confidence interval for β_1 is [______, ______]

(*Round your response to two decimal places*)

Supposed you learned that Y_i and X_i were independent. Would you be surpised?

- **A.** Yes, I would be surprised because the null hypothesis that β_1 is zero was not rejected at the 5% significance level.
- **B.** No, I wouldn't be surprised because the null hypothesis that β_1 is zero was not rejected at the 5% significance level.
- **C.** No, I wouldn't be surprised because the null hypothesis that β_1 is zero was rejected at the 5% significance level.
- **D.** Yes, I would be surprised because the null hypothesis that β_1 is zero was rejected at the 5% significance level.

Suppose that Y_i and X_i are independent and many samples of size n = 257 are drawn and regressions estimated. Suppose that you test the null hypothesis that β_1 is zero at the 5% level and construct a 95% confidence interval for β_1.

In what fraction of the samples would the null hypothesis that β_1 is zero at the 5% level be rejected?

- **A.** The null hypothesis would be rejected in 95% of the samples.
- **B.** The null hypothesis would be rejected in 5% of the samples.
- **C.** The null hypothesis would be rejected in 10% of the samples.
- **D.** The null hypothesis would be rejected in 90% of the samples.

In what fraction of the samples would the value $\beta_1 = 0$ be included in the 95% confidence interval for β_1?

- **A.** 5% of the confidence intervals would contain the value $\beta_1 = 0$.
- **B.** 10% of the confidence intervals would contain the value $\beta_1 = 0$.

- **C.** 95% of the confidence intervals would contain the value $\beta_1 = 0$.
- **D.** 90% of the confidence intervals would contain the value $\beta_1 = 0$.

Answers 3.31

0.00

A. Yes

2.30

8.96

D. Yes, I would be surprised because the null hypothesis that β_1 is zero was rejected at the 5% significance level.

B. The null hypothesis would be rejected in 5% of the samples.

C. 95% of the confidence intervals would contain the value $\beta_1 = 0$.

ID: Exercise 5.7

---

15. The 95% confidence interval for β_1 is the interval:

- **A.** $\left(\beta_1 - 1.96SE\left(\beta_1\right), \beta_1 + 1.96SE\left(\beta_1\right)\right)$.
- **B.** $\left(\hat{\beta}_1 - 1.645SE\left(\hat{\beta}_1\right), \hat{\beta}_1 + 1.645SE\left(\hat{\beta}_1\right)\right)$.
- **C.** $\left(\hat{\beta}_1 - 1.96, \hat{\beta}_1 + 1.96\right)$.
- **D.** $\left(\hat{\beta}_1 - 1.96SE\left(\hat{\beta}_1\right), \hat{\beta}_1 + 1.96SE\left(\hat{\beta}_1\right)\right)$.

Answer: D. $\left(\hat{\beta}_1 - 1.96SE\left(\hat{\beta}_1\right), \hat{\beta}_1 + 1.96SE\left(\hat{\beta}_1\right)\right)$.

ID: Test A Ex 5.2.3

---

16. Using 143 observations, assume that you had estimated a simple regression function and that your estimate for the slope was 0.04, with a standard error of 0.01. You want to test whether or not the estimate is statistically significant.

Which of the following decisions is the only correct one?

- **A.** Since the slope is very small, so must be the regression R^2.
- **B.** You decide that the coefficient is small and hence most likely is zero in the population.
- **C.** The response of *Y* given a change in *X* must be economically important since it is statistically significant.
- **D.** The slope is statistically significant since it is four standard errors away from zero.

Answer: D. The slope is statistically significant since it is four standard errors away from zero.

ID: Test A Ex 5.2.4

---

17. You have collected data for the 50 U.S. states and estimated the following relationship between the change in the unemployment rate from the previous year ($\widehat{\triangle ur}$) and the growth rate of the respective state real GDP (g_y). The results are as follows:

$$\widehat{\triangle ur} = 2.81 - 0.23 \times g_y,\ R^2 = 0.36,\ SER = 0.78$$
$$(0.12)\ (0.04)$$

Assuming that the estimator has a normal distribution, the 95% confidence interval for the slope is approximately the interval:

- **A.** $[-0.31, 0.15]$.
- **B.** $[-0.31, -0.15]$.
- **C.** $[-0.33, -0.13]$.
- **D.** $[2.57, 3.05]$.

Answer: B. $[-0.31, -0.15]$.

ID: Test B Ex 5.2.2

---

18. You extract approximately 5,000 observations from the Current Population Survey (CPS) and estimate the following regression function:

$$\widehat{AHE} = 3.32 - 0.45 \times Age,\ R^2 = 0.02,\ SER = 8.66$$
$$(1.00)\ (0.04)$$

where *AHE* is average hourly earnings, and *Age* is the individual's age.

Given the specification, your 95% confidence interval for the effect of changing age by 5 years is approximately:

- **A.** [$1.86, $2.64].
- **B.** [$1.35, $5.30].
- **C.** [$2.32, $4.32].
- **D.** Cannot be determined given the information provided.

Answer: A. [$1.86, $2.64].

ID: Test B Ex 5.2.3

---

19. Suppose that a researcher, using data on 150 randomly selected bicycles, estimates the OLS regression:

$$\underset{(191.4)}{\widehat{Price}} = 650.8 - \underset{(0.42)}{1.98} \times Weight,$$

where *Price* measures the price of the i^{th} bike in dollars and *Weight* measures the weight of the i^{th} bike in kilograms.

The 99% confidence interval for the intercept, β_0, will be (____ , ____).

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

The 99% confidence interval for the slope, β_1, will be (____ , ____).

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

Based on the calculated confidence intervals, and a two-tailed hypothesis test, we can say that at the 1% significance level, we will (1) ______ the hypothesis $\beta_0 = 165$, and we will (2) ______ the hypothesis $\beta_1 = -1$.

(1) ○ fail to reject
○ reject

(2) ○ reject
○ fail to reject

Answers 157.95

1,143.66

-3.06

-0.90

(1) fail to reject

(2) fail to reject

ID: Concept Exercise 5.2.1

20. A professor wants to understand the relationship between students' class attendance (X_i) and the grades that the students secure in the final exam (Y_i). The professor selects 108 students at random for the experiment.

The estimated OLS regression is:

$$\hat{Y}_i = 49.87 + 1.86X_i,$$

where $\hat{Y}_i$ denotes the predicted value of the grades obtained by the i^{th} student and X_i denotes the number of classes the student attends.

Let $\beta_1 \Delta x$ represent the predicted change in the test score associated with a small change in the attendance (Δx). Suppose that the professor wants to see the effect of a fall in attendance by 2 days on the test scores.

If the standard error of the estimated slope $\hat{\beta}_1$ is 0.85, the 95% confidence interval for $\beta_1 \Delta x$ will be: [___ , ___].

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

A 95% confidence interval for the effect of increasing the attendance by 5 days on test scores could be as great as ___ or as little as ___.

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

Answers − 0.39

− 7.05

17.63

0.97

ID: Concept Exercise 5.2.2

21. Suppose that a researcher, using wage data on 235 randomly selected male workers and 263 female workers, estimates the OLS regression

$$\widehat{Wage} = \underset{(0.2162)}{11.769} + \underset{(0.3384)}{1.993} \times Male, \; R^2 = 0.04, \; SER = 3.9,$$

where *Wage* is measured in dollars per hour and *Male* is a binary variable that is equal to 1 if the person is a male and 0 if the person is a female. Define the wage gender gap as the difference in mean earnings between men and women.

What is the estimated gender gap?

The estimated gender gap equals $[ ] per hour. (*Round your response to three decimal places.*)

The null and alternative hypotheses are $H_0: \hat{\beta}_1 = 0$ versus $H_1: \hat{\beta}_1 \neq 0$.

The *t*-statistic for testing the null hypothesis that there is no gender gap is [ ]. (*Round your response to two decimal places.*)

The *p*-value for testing the null hypothesis that there is no gender gap is [ ]. (*Round your response to four decimal places.*)

The estimated effect of gender gap is statistically significant at the:

I. 5% level
II. 1% level
III. 0.01% level

- **A.** I and II.
- **B.** I, II, and III.
- **C.** I only.
- **D.** III only.

Construct a 95% confidence interval for the effect of gender gap.

The 95% confidence interval for the effect of gender gap is ([ ], [ ]). (*Round your responses to two decimal places.*)

From the sample, the average wage of women is $[ ] per hour. (*Round your response to three decimal places.*)

From the sample, the average wage of men is $[ ] per hour. (*Round your response to three decimal places.*)

Another researcher uses these same data but regresses *Wages* on *Female*, a variable that is equal to 1 if the person is female and 0 if the person a male. What are the regression estimates calculated from this regression?

$\overline{\text{Wage}} = \hat{\gamma}_0 + \hat{\gamma}_1 \times Female, \; R^2, \; SER.$

$\hat{\gamma}_0$ = [ ]. (*Round your response to three decimal places.*)

$\hat{\gamma}_1$ = [ ]. (*Round your response to three decimal places.*)

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

SER = [ ]. (*Round your response to one decimal place.*)

Answers 1.993

5.89

0.0000

B. I, II, and III.

1.33

2.66

11.769

13.762

13.762

– 1.993

0.04

3.9

ID: Exercise 5.2

22. In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to "regular" and "small" classes, and given standardized tests at the end of the year. (Regular classes contained approximately 24 students, and small classes contained approximately 15 students.) Suppose that, in the population, the standardized tests have a mean score of 897 points and a standard deviation of 73 points. Let *SmallClass* denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise.

A regression of *TestScore* on *SmallClass* yields

$$\widehat{TestScore} = \underset{(1.7)}{890.5} + \underset{(2.7)}{15.2} \times SmallClass, \; R^2 = 0.01, \; SER = 72.4.$$

Do small classes improve test scores? By how much? Is the effect large?

The estimated gain from being in a small class is [ ] points, (1) ________. (*Round your response to one decimal place.*)

The null and alternative hypotheses are: $H_0: \hat{\beta}_1 = 0$ versus $H_1: \hat{\beta}_1 \neq 0$.

The *t*-statistic for testing the null hypothesis that small classes do not improve test scores is [ ]. (*Round your response to two decimal places.*)

The *p*-value for testing the null hypothesis that small classes do not improve test scores is [ ]. (*Round your response to four decimal places.*)

The estimated effect of class size on test scores is statistically significant at the:

I. 10% level
II. 5% level
III. 1% level
IV. 0.5% level
V. 0.1% level

- **A.** V only.
- **B.** I, II, and III.
- **C.** I only.
- **D.** I, II, III, IV, and V.

Construct a confidence interval of 90% for the effect of *SmallClass* on test score.

The 90% confidence interval for the effect of *SmallClass* on test score is ([ ], [ ]). (*Round your responses to one decimal place.*)

(1)
- a significant increase
- a moderate increase
- an insignificant increase

Answers 15.2

(1) a moderate increase

5.63

0.0000

D. I, II, III, IV, and V.

10.8

19.6

ID: Exercise 5.5

23. Consider the regression model

$$Y_i = \beta X_i + u_i$$

Where u_i and X_i satisfy the assumptions specified here. Let $\bar{\beta}$ denote an estimator of β that is constructed as $\bar{\beta} = \frac{\bar{Y}}{\bar{X}}$, where $\bar{Y}$ and $\bar{X}$ are the sample means of Y_i and X_i, respectively.

Show that $\bar{\beta}$ is a linear function of $Y_1, Y_2, \ldots, Y_n$.

$$\bar{\beta} = \frac{\bar{Y}}{\bar{X}} = \frac{(1)\ \_\_\_\_\_\_\_\_}{\bar{X}}$$

Show that $\bar{\beta}$ is conditionally unbiased.

1. $E\left(Y_i | X_1, X_2, \ldots, X_n\right)$ = (2) ________

2. $E\left(\bar{\beta} | X_1, X_2, \ldots, X_n\right) = E\left[\left(\frac{\frac{1}{n}\left(Y_1 + Y_2 + \ldots + Y_n\right)}{\bar{X}}\right) | \left(X_1, X_2, \ldots, X_n\right)\right] = \frac{(3)\ \_\_\_\_\_\_\_\_}{\bar{X}} = \beta$

(1)
- ○ $\frac{1}{n}\beta\left(Y_1 + Y_2 + \ldots + Y_n\right)$
- ○ $\frac{1}{n}\beta\left(X_1 + X_2 + \ldots + X_n\right)$
- ○ $\frac{1}{n}\left(\beta X_i + u_i\right)$

(2)
- ○ βX_i
- ○ $\beta \bar{X}$
- ○ $\beta \bar{Y}$

(3)
- ○ $\frac{1}{n}\beta\left(Y_1 + Y_2 + \ldots + Y_n\right)$
- ○ $\frac{1}{n}\beta\left(X_1 + X_2 + \ldots + X_n\right)$
- ○ $\frac{1}{n}\left(\beta X_i + u_i\right)$

Answers (3) $\frac{1}{n}\beta\left(X_1 + X_2 + \ldots + X_n\right)$

(2) βX_i

(3) $\frac{1}{n}\beta\left(X_1 + X_2 + \ldots + X_n\right)$

ID: Exercise 5.9

24. Let X_i denote a binary variable and consider the regression $Y_i = \beta_0 + \beta_1 X_i + u_i$. Let $\bar{Y}_0$ denote the sample mean for observations for X = 0 and $\bar{Y}_1$ denote the sample mean for observations with X = 1. Show that $\hat{\beta}_0 = \bar{Y}_0$, $\hat{\beta}_0 + \hat{\beta}_1 = \bar{Y}_1$, and $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$.

Let n_0 denote the number of observations with X = 0 and n_1 denote the number of observations with X = 1.

Show that $\bar{\beta}$ is conditionally unbiased.

1. $E\left(Y_i | X_1, X_2, \ldots, X_n\right)$ = (1) ____________

2. $E\left(\bar{\beta} | X_1, X_2, \ldots, X_n\right) = E\left[\left[\left(\dfrac{\frac{1}{n}\left(Y_1 + Y_2 + \ldots + Y_n\right)}{\bar{Y}_1}\right) \Big| \left(X_1, X_2, \ldots, X_n\right)\right]\right] = \dfrac{(2)\ \_\_\_\_\_\_\_\_\_\_\_\_}{\bar{Y}_1} = \beta$

(1)
- ○ βX_i
- ○ $\beta \bar{Y}_1$
- ○ $\beta \bar{Y}_0$

(2)
- ○ $\frac{1}{n}\beta\left(Y_1 + Y_2 + \ldots + Y_n\right)$
- ○ $\frac{1}{n}\beta\left(X_1 + X_2 + \ldots + X_n\right)$
- ○ $\frac{1}{n}\left(\beta X_i + u_i\right)$

Answers (1) βX_i

(2) $\frac{1}{n}\beta\left(X_1 + X_2 + \ldots + X_n\right)$

ID: Exercise 5.10

25. 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 [1]. Use a statistical package of your choice to answer the following questions.

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

Is the estimated slope statistically significant?

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

Construct a 95% confidence interval for the slope coefficient using heteroskedasticity–robust standard errors [2].

The 95% confidence interval for the slope coefficient is [,]

(*Round your responses to three decimal places*)

Run a regression of *Earnings* on *Height* using data for female workers only.

Is the estimated slope statistically significant?

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

Construct a 95% confidence interval for the slope coefficient using heteroskedasticity-robust standard errors [3].

The 95% confidence interval for the slope coefficient is [,]

(*Round your responses to three decimal places*)

Run a regression of *Earnings* on *Height* using data for male workers only.

Is the estimated slope statistically significant?

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

Construct a 95% confidence interval for the slope coefficient using heteroskedasticity-robust standard errors [4].

The 95% confidence interval for the slope coefficient is [,]

(*Round your responses to three decimal places*)

Can you reject the null hypothesis that the effect of height on earnings is the same for men and women?

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

1: More Info

Variable Definitions

Variable	Definition
Earnings	Annual labor earnings, expressed in $2012.
Height	Height without shoes (in inches).
Sex	1 = Male; 0 = Female

2: More Info

Heteroskedasticity–Robust Standard Errors

Heteroskedasticity-robust standard errors are calculated as follows:

$$SE\left(\hat{\beta}_1\right) = \sqrt{\hat{\sigma}^2_{\beta_1}}$$

Where

$$\hat{\sigma}^2_{\beta_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{t=1}^{n}\left(X_i - \bar{X}\right)^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum^{n}\left(X_i - \bar{X}\right)^2\right]^2}$$

3: More Info

Heteroskedasticity–Robust Standard Errors

Heteroskedasticity-robust standard errors are calculated as follows:

$$SE\left(\hat{\beta}_1\right) = \sqrt{\hat{\sigma}^2_{\beta_1}}$$

Where

$$\hat{\sigma}^2_{\beta_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{t=1}^{n}\left(X_i - \bar{X}\right)^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2}$$

4: More Info

Heteroskedasticity–Robust Standard Errors

Heteroskedasticity-robust standard errors are calculated as follows:

$$SE\left(\hat{\beta}_1\right) = \sqrt{\hat{\sigma}^2_{\beta_1}}$$

Where

$$\hat{\sigma}^2_{\beta_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{t=1}^{n}\left(X_i - \bar{X}\right)^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2}$$

Answers B. No.

– 928.378

2147.964

B. No.

– 10925.157

5052.651

B. No.

– 8592.239

22605.405

B. No.

ID: Empirical Exercise 5.1

26. 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 data set is available here [5]. Use a statistical package of your choice to answer the following questions.

Run a regression of *Growth* on *TradeShare*.

Test the null hypothesis $H_0: \beta_1 = 0$ versus a two-sided alternative hypothesis $H_1: \beta_1 \neq 0$. Compute the *t*-statistic and *p*-value.

The *t*-statistic is ______.

(*Round your response to three decimal places*)

The *p*-value is ______.

(*Round your response to three decimal places*)

Is the estimated regression slope statistically significant? That is, can you reject the null hypothesis $H_0: \beta_1 = 0$ versus a two-sided alternative hypothesis $H_1: \beta_1 \neq 0$ at the 10% significance level?

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

Construct a 90% confidence interval for β_1.

The 90% confidence interval is [______, ______]

(*Round your response to three decimal places*)

5: More Info

Variable Definitions

Variable	Definition
Growth	Average annual percentage growth of real Gross Domestic Product (GDP) from 1960 to 1995.
Tradeshare	The average share of trade in the economy from 1960 to 1995, measured as the sum of exports plus imports, divided by GDP; that is, the average value of $\frac{(X+M)}{GDP}$ from 1960 to 1995, where X = exports and M = imports (both X and M are positive).

Answers 1.703

0.089

A. Yes.

0.052

3.310

ID: Empirical Exercise 5.2

27. In this exercise you will investigate the relationship between birth weight and smoking during pregnancy.

The following table contains data for a random sample of babies born in Pennsylvania in 1989. The data includes the baby's birth weight together with various characteristics of the mother, including whether she smoked during her pregnancy. Download the data from the table by clicking the *download table* icon . A detailed description of the variables used in the data set is available here [6]. Use a statistical package of your choice to answer the following questions.

What is the average value of *Birthweight* for all mothers in the sample?

The average value of *Birthweight* for all mothers in the sample is ____ grams.

(*Round your response to three decimal places*)

What is the average value of *Birthweight* for mothers who smoke?

The average value of *Birthweight* for mothers who smoke is ____ grams.

(*Round your response to three decimal places*)

What is the average value of *Birthweight* for mothers who do not smoke?

The average value of *Birthweight* for mothers who do not smoke is ____ grams.

(*Round your response to three decimal places*)

Use the data in the sample to estimate the difference in average birth weight for smoking and nonsmoking mothers.

The difference in average birth weight for smoking and nonsmoking mothers is ____ grams.

(*Round your response to three decimal places*)

What is the standard error for the estimated difference?

The standard error for the estimated difference is ____ grams.

(*Round your response to three decimal places*)

Construct a 95% confidence interval for the difference in the average birth weight for smoking and nonsmoking mothers.

The 95% confidence interval for the difference in the average birth weight for smoking and nonsmoking mothers is [____ , ____] grams.

(*Round your response to three decimal places*)

Run a regression of *Birthweight* on the binary variable *Smoker*.

Which of the following is true about the estimated slope and intercept? (*Check all that apply*)

- ☐ **A.** The estimated intercept is the average birth weight for nonsmoking mothers.
- ☐ **B.** The estimated slope is the expected decrease in birth weight for every additional cigarette a mother smokes.
- ☐ **C.** The estimated intercept plus the estimated slope is the average birth weight for smoking mothers.
- ☐ **D.** The estimated slope is the difference in average birth weight for smoking and nonsmoking mothers.

Explain how the $SE\left(\hat{\beta}_1\right)$ is related to the standard error of the estimated difference in average birth weight for smoking and nonsmoking mothers.

- ○ **A.** $SE\left(\hat{\beta}_1\right)$ is equal to the standard error of the estimated difference in average birth weights for smoking and nonsmoking mothers if we assume homoskedasticity.
- ○ **B.** $SE\left(\hat{\beta}_1\right)$ is greater than the standard error of the estimated difference in average birth weights for smoking and nonsmoking mothers.

C. $SE(\hat{\beta}_1)$ is less than the standard error of the estimated difference in average birth weights for smoking and nonsmoking mothers.

D. $SE(\hat{\beta}_1)$ is equal to the standard error of the estimated difference in average birth weights for smoking and nonsmoking mothers if we assume heteroskedasticity.

Construct a 95% confidence interval for the effect of smoking on birth weight using heteroskedasticity–robust standard errors.

The 95% confidence interval for the effect of smoking on birth weight is [,] grams.

(*Round your response to three decimal places*)

Do you think that smoking is uncorrelated with other factors that cause low birth weight? That is, do you think that the regression error term has a conditional mean zero, given *Smoking* (X_i)?

A. Yes.

B. No.

6: More Info

Variable Definitions

Variable	Definition
BirthWeight	Birth weight of infant (in grams).
Smoker	Indicator; = 1 if the mother smoked during pregnancy and 0 otherwise.

Answers 3248.570

3098.800

3398.340

299.540

116.780

70.651

528.429

A. The estimated intercept is the average birth weight for nonsmoking mothers., C.
The estimated intercept plus the estimated slope is the average birth weight for smoking mothers., D.
The estimated slope is the difference in average birth weight for smoking and nonsmoking mothers.

D.
$SE(\hat{\beta}_1)$ is equal to the standard error of the estimated difference in average birth weights for smoking and nonsmoking mother
we assume heteroskedasticity.

– 528.428

– 70.652

B. No.

ID: Empirical Exercise 5.3

28. 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 data set is available here [7]. 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*)

The estimated model is

$$\widehat{Ahe} = -2.1843 + \underset{(0.396)}{0.7458} Age, \quad R^2 = 0.035$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Suppose you wanted to test the hypothesis that β_1 equals zero at the 1%, 5% and 10% level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is ______.

(*Round your response to three decimal places*)

The *p*-value is ______.

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

Construct a 95% confidence interval for the slope coefficient β_1.

The 95% confidence interval for the slope coefficient β_1 is [______, ______]

(*Round your response to three decimal places*)

Now, suppose you are interested in estimating the following model for the subpopulation of high school graduates (*college = 0*)

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

Run a regression of average hourly earnings (*AHE*) on age (*Age*) using the sample for which the indicator variable *college* = 0.

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*)

The estimated model is

$$\underset{}{Ahe} = -12.8098 + \underset{(0.374)}{0.9827} Age,\ R^2 = 0.110$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Using the sample of high school graduates, suppose you wanted to test the hypothesis that β_1 equals zero at the 5% significance level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is [ ].

(*Round your response to three decimal places*)

The *p*-value is [ ].

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

Now, suppose you are interested in estimating the following model for the subpopulation of college graduates (*college = 1*)

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

Run a regression of average hourly earnings (*AHE*) on age (*Age*) using the sample for which the indicator variable *college* = 1.

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*)

The estimated model is

$$Ahe = 28.3547 + \underset{(0.678)}{-0.1080} Age,\ R^2 = 0.001$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Using the sample of college graduates, suppose you wanted to test the hypothesis that β_1 equals zero at the 5% significance level. That is

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is ______.

(*Round your response to three decimal places*)

The *p*-value is ______.

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

Is the effect of age on earnings different from high school graduates that for college graduates? In other words, can we reject the following hypothesis at a reasonable significance level?

$$H_0: \beta_1^{\text{college}} - \beta_1^{\text{high school}} = 0 \text{ vs. } H_1: \beta_1^{\text{college}} - \beta_1^{\text{high school}} \neq 0$$

Use the following formula to report the *t*-statistic and *p*-value for this test.

$$SE\left(\hat{\beta}_1^{\text{college}} - \hat{\beta}_1^{\text{high school}}\right) = \sqrt{\left(\left(SE_{\hat{\beta}_1,\text{ college}}\right)^2 + \left(SE_{\hat{\beta}_1,\text{ high school}}\right)^2\right)}$$

The *t*-statistic is ______.

(*Round your response to three decimal places*)

The *p*-value is ______.

(*Round your response to three decimal places*)

7: More Info

Variable Definitions

Variable	**Definition**
Ahe	Average hourly earnings.
Age	Age in years.
College	= 1 if worker has a bachelor's degree; = 0 if worker has a high school degree.

Answers -2.1843

0.7458

1.883

0.060

B. No.

-0.031

1.522

-12.8098

0.9827

2.627

0.009

B. Yes.

28.3547

-0.1080

-0.159

0.874

A. No.

-1.408

0.159

ID: General Empirical 5.1

29. Using the Excel data set CPS08[8] (described[9]), run a regression of average hourly earnings (*AHE*) on *age* and answer the following questions.

The coefficient on *age* shows the

- A. the change in age for every $5.00 increase in average hourly earnings
- B. the change in hourly earnings for every 5-year increase in age
- C. the change in hourly earnings for every 1-year increase in age
- D. the change in age for every $1.00 increase in average hourly earnings

Given the following hypothesis: H_0: β_{age} = 0 where β_{age} is the coefficient on *age* you would:

- A. None of the above
- B. Not reject H_0 and because the coefficient on age is so small in magnitude
- C. Reject H_0 at the 5% level because the t-statistic is greater than 1.96
- D. Not reject H_0 and conclude that the coefficient on age is statistically insignificant at the 5% level.

The 95% confidence interval associated with the null hypothesis: H_0: β_{age} =0 is

- A. -1.969 - 4.133
- B. 0.204 - 0.489
- C. 0.527 - 0. 683
- D. 0.039 - 0.605

Given the following hypothesis: H_0: β_{age} =0.50 where β_{age} is the coefficient on Age you would:

- A. Reject H_0 because the t-statistic associated with the null is 2.63
- B. Not reject H_0 because the estimated coefficient on *age* is close to 0.50
- C. Not reject H_0 because the 95% confidence interval includes 0.50
- D. There is not enough information to test the null hypothesis

Based on the regression the expected average hourly wage for a person 20 years of age would be:

- A. Would be $1.32 which shows how poorly age explains earnings
- B. Would be $21.65 which is obtained by multiplying the intercept by 20
- C. Would be $13.18 which is plausible for a 20 year old
- D. Would be $14.82 which is more than a person 25 years of age would earn.

Re-run the regression in (1) but estimate it only for women (Hint: the number of observations in the regression should be 3,336). The results indicate that

- A. the average hourly earning of a woman 30 years of age is $3.46 more than a woman 20 years of age
- B. the coefficient on age is 0.346 but it is not statistically significant at the 95% confidence level
- C. the average hourly earnings of women is $7.27 less than males
- D. none of the answers in (a)-(c) are correct.

Re-run the regression in (1) but now estimate it only for men (Hint: the number of observations in the regression should be 4,375). The results indicate that

- **A.** the expected average hourly earnings for a man 30 years of age is $20.38
- **B.** men actually earn $3.05 less per hour than women
- **C.** the coefficient on age is 0.78 but it is not statistically significant at the 95% confidence level
- **D.** every year of age increases a man's hourly earning by $7.81

Based on the separate regressions for men and women we would conclude that

- **A.** the negative intercept in the regression for males means they earn $3.05 less than women as age increases
- **B.** the expected hourly earnings of men increases by $0.43 more than women for every one year increase in age
- **C.** the expected average hourly earnings for a man 30 years of age is $20.38 whereas the expected hourly earnings for woman of the same age is $15.66
- **D.** women earn slightly more than men for every year of age

(Challenging question) In Exercise 5.15, Stock and Watson describe a means by which one could statistically test whether the coefficient o *age* in the female regression is different from the coefficient on *age* in the male regression. Following their approach, the 95% confidence fo the difference of the two coefficients would be

- **A.** 0.238 - 0454
- **B.** 0.588 - 0.281
- **C.** 0.672 - 0.890
- **D.** 0.112 - 0.740

(Challenging question) Using the results in part (9), test the following null hypothesis,H_0: $\beta_{age,male} - \beta_{age, female} = 0$.

- **A.** Reject the null because the 95% confidence interval includes zero
- **B.** Do not reject the null
- **C.** Reject the null because the 95% confidence interval does not include zero
- **D.** Do not reject the null because the difference in the two coefficients is only 0.433

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

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

Answers C. the change in hourly earnings for every 1-year increase in age

C. Reject H_0 at the 5% level because the t-statistic is greater than 1.96

C. 0.527 - 0. 683

A. Reject H_0 because the t-statistic associated with the null is 2.63

C. Would be $13.18 which is plausible for a 20 year old

A. the average hourly earning of a woman 30 years of age is $3.46 more than a woman 20 years of age

A. the expected average hourly earnings for a man 30 years of age is $20.38

B. the expected hourly earnings of men increases by $0.43 more than women for every one year increase in age

B. 0.588 - 0.281

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

30. 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 . Use a statistical package of your choice to answer the following questions.

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*)

The estimated model is

$$\widehat{\textit{Course Evaluation}} = \underset{}{3.920} + \underset{(0.088)}{0.028}\textit{Age}, \quad R^2 = 0.001$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Suppose you wanted to test the hypothesis that β_1 equals zero at the 1%, 5% and 10% level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is ______.

(*Round your response to two decimal places*)

The *p*-value is ______.

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

Answers 3.920

0.028

0.32

0.749

A. No.

ID: General Empirical 5.2

31. 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 [10]. 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 completed of education (*ED*) on distance to the nearest college (*Dist*).

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*)

The estimated model is

$$\widehat{ED} = \underset{}{13.534} + \underset{(0.068)}{-0.010} Dist, \quad R^2 = 0.000$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Suppose you wanted to test the hypothesis that β_1 equals zero at the 1%, 5% and 10% level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is ______.

(*Round your response to two decimal places*)

The *p*-value is ______.

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

Construct a 95% confidence interval for the slope coefficient β_1.

The 95% confidence interval for the slope coefficient β_1 is [______, ______]

(*Round your response to three decimal places*)

Now, suppose you are interested in estimating the following model for the subpopulation of females (*female* = 1)

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

Run a regression of years completed of education (*ED*) on distance to the nearest college (*Dist*) using the sample for which the indicator variable *female* = 1.

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*)

The estimated model is

$$\widehat{ED} = \underset{}{13.878} - \underset{(0.121)}{0.088} Dist, \quad R^2 = 0.011$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Suppose you wanted to test the hypothesis that β_1 equals zero at the 1%, 5% and 10% level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is [ ].

(*Round your response to two decimal places*)

The *p*-value is [ ].

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

Now, suppose you are interested in estimating the following model for the subpopulation of males (*female* = 0)

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

Run a regression of years completed of education (*ED*) on distance to the nearest college (*Dist*) using the sample for which the indicator variable *female* = 0.

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*)

The estimated model is

$$\widehat{ED} = \underset{}{13.256} + \underset{(0.080)}{0.038} Dist, \quad R^2 = 0.005$$

Where the number in parentheses is the homoskedastic standard error for the regression coefficient $\hat{\beta}_1$.

Suppose you wanted to test the hypothesis that β_1 equals zero at the 1%, 5% and 10% level. That is,

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

Report the *t*-statistic and *p*-value for this test.

The *t*-statistic is [ ].

(*Round your response to two decimal places*)

The *p*-value is [ ].

(*Round your response to three decimal places*)

Can you reject the null hypothesis at the 5% significance level?

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

10: More Info

Variable Definitions

Variable	Definition
Education	Years of Education completed.
Distance	Distance from 4 year college in 10's of miles.
Female	Indicator variable. = 1 if female and = 0 if male.

Answers 13.534

− 0.010

− 0.15

0.881

B. No.

− 0.144

0.124

13.878

− 0.088

− 0.72

0.472

B. No.

13.256

0.038

0.48

0.631

A. No.

ID: General Empirical 5.3

32. Using the Excel data set, CollegeDistance[11], described[12] in Empirical Exercise 4.3, run a regression of years of completed schooling (*ed*) on distance (in 10s of miles) from a 4-year college (*dist*).

The coefficient on distance (*dist*) shows the

- **A.** the change in completed schooling for every 10 mile increase in distance to the nearest college
- **B.** the change in distance for every year of complete schooling
- **C.** the change in completed schooling for every 1 mile increase in distance to the nearest college
- **D.** the change in distance for every 10 years of completed schooling

Given the following hypothesis: H_0: β_{dist} =0 where β_{dist} is the coefficient on distance you would:

- **A.** Not reject H_0 because the t-ratio, -5.33, is less than 1.96
- **B.** None of the above
- **C.** Not reject H_0 at the 5% level because the 95% confidence interval does not include zero
- **D.** Reject H_0 because the t-ratio, -5.33, is greater than 1.96 in absolute value

The 99% confidence interval associated with the null hypothesis: H_0: β_{dist} = 0 is

- **A.** -0.100 - -0.046
- **B.** 0.038 - 0. 109
- **C.** -0.109 - -0.038
- **D.** -0.073 - 0.104

Given the following hypothesis: H_0: β_{dist} = -0.150 where β_{dist} is the coefficient on distance you would:

- **A.** Reject H_0 because the t-statistic associated with the null is 5.57
- **B.** There is not enough information to test the null hypothesis
- **C.** Not reject H_0 because the estimated t-statistic is less than 2 in absolute value
- **D.** Not reject H_0 because the 95% confidence interval includes -0.150

Based on the regression the expected years of completed schooling for a person 100 miles from the nearest 4-year college is:

- **A.** None of the above
- **B.** 13.88
- **C.** 13.22
- **D.** 6.62

Re-run the regression in (1) but estimate it only for females (Hint: the number of observations in the regression should be 2,070). The results indicate that

- **A.** the coefficient on distance is -0.064 and it is not statistically significant at the 95 % confidence level because the t-ratio is less than zero
- **B.** the coefficient on distance is -0.064 and there is not enough information to test its statistical significance at the 99% confidence level.
- **C.** the coefficient on distance is -0.064 but it is not statistically significant at the 95% confidence level because the confidence interval includes zero
- **D.** the coefficient on distance is -0.064 but it is statistically significant at the 99% confidence level because the prob-value is less than 0.01

Re-run the regression in (1) but now estimate it only for males (Hint: the number of observations in the regression should be 1,726). The results indicate that the expected years of schooling for

- **A.** a man who lived 100 miles from a 4-year college are 13.1
- **B.** a man who lived 10 miles from a 4-year college are 8.3
- **C.** a man who lived 200 miles from a 4-year college are 12.3
- **D.** both (b) and (c) are correct

Based on the separate regressions for men and women we would conclude that

- **A.** every 10 miles from a 4-year colleges lowers completed schooling more for men than for women
- **B.** the expected years of completed schooling associated with living 100 miles from a 4-year college is 13.3 for males and 13.1 for females
- **C.** every 10 miles from a 4-year colleges lowers completed schooling more for women than for men
- **D.** None of the above

(Challenging question) In Exercise 5.15, Stock and Watson describe a means by which one can statistically test whether the coefficient on distance in the female regression is different from the coefficient on distance in the male regression. Following their approach, the estimated t-statistic (or z-statistic given the large number of observations) associated with difference in the two slopes coefficients is

- **A.** 0.71
- **B.** 7.13
- **C.** -2.96
- **D.** 2.96

(Challenging question. Use the results from part 9) Given the following null hypothesis, H_0: $\beta_{dist,male} - \beta_{dist, female} = 0$, we would

- **A.** Not reject the null because the associated t-statistic is less than 1.96 in absolute value
- **B.** Not reject the null because the difference in the two coefficients is only -0.0197
- **C.** Reject the null at the 0.05 level because the t-statistic is 1.96
- **D.** Reject the null at the 0.05 level because the t-statistic is less than -1.96

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

12: http://https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_3/empirical/empex_tb/CollegeDistance_DataDescript

Answers A. the change in completed schooling for every 10 mile increase in distance to the nearest college

D. Reject H_0 because the t-ratio, -5.33, is greater than 1.96 in absolute value

C. -0.109 - -0.038

A. Reject H_0 because the t-statistic associated with the null is 5.57

C. 13.22

D.
the coefficient on distance is -0.064 but it is statistically significant at the 99% confidence level because the prob-value is less thá
0.01

D. both (b) and (c) are correct

A. every 10 miles from a 4-year colleges lowers completed schooling more for men than for women

A. 0.71

A. Not reject the null because the associated t-statistic is less than 1.96 in absolute value

ID: General Empirical 5.3 (static)

---

33. A binary variable is often called a:

- **A.** power of a test.
- **B.** residual.
- **C.** dependent variable.
- **D.** dummy variable.

Answer: D. dummy variable.

ID: Test B Ex 5.3.4

---

34. A study tests the effect of earning a Master's degree on the salaries of professionals. Suppose that the salaries of the professionals (S_i) are not dependent on any other variables. Let D_i be a variable which takes the value 0 if an individual has not earned a Master's degree, and a value 1 if they have earned a Master's degree.

What would be the regression model that the researcher wants to test?

- **A.** $0 = \beta_0 + \beta_1 S_i + u_i$, $i = 1, \ldots, n$.
- **B.** $S_i = \beta_0 + \beta_1 D_i + u_i$, $i = 1, \ldots, n$.
- **C.** $1 = \beta_0 + \beta_1 S_i + u_i$, $i = 1, \ldots, n$.
- **D.** $S_i = \beta_0 + \beta_1 + u_i$, $i = 1, \ldots, n$.

Suppose that a random sample of 150 individuals suggests that professionals without a Master's degree earn an average salary of $58,000 per annum, while those with a Master's degree earn an average salary of $80,000 per annum.

The OLS estimate of the coefficient β_1 will be $[ ] and that of β_0 will be $[ ].

Answers B. $S_i = \beta_0 + \beta_1 D_i + u_i$, $i = 1, \ldots, n$.

22,000

58,000

ID: Concept Exercise 5.3.1

35. After studying the milk yields of a randomly selected sample of 900 cows, researchers felt that the presence of white spots (WS_i) on a cow's body led to a different level of milk production (MP_i). The researchers estimate the OLS regression:

$$\widehat{MP}_i = 5.84 + 0.14 \times WS_i,$$
$$(2.23) \quad (0.06)$$

where WS_i is a binary variable of the form:

$$WS_i = \begin{cases} 1, \text{ if the } i^{th} \text{ cow has white spots} \\ 0, \text{ if the } i^{th} \text{ cow doesn't have white spots} \end{cases}.$$

The 95% confidence interval for the coefficient β_1 is (____ , ____).

(*Round your answers to four decimal places.*)

Based on this interval, we will (1) ____ the hypothesis $\beta_1 = 0$ at the 5% significance level.

Which of the following statements describes the way in which the coefficient of the indicator variable could be interpreted in this case?

- **A.** The coefficient is the sum of the conditional expectation of the presence of white spots on a cow with respect to the average milk production.
- **B.** The coefficient is the rate of change in the milk production due to a unit change in the number of cows with white spots.
- **C.** The coefficient of the indicator variable is the slope of the regression equation.
- **D.** The coefficient is the difference between the population means of the milk production of cows with and without white spots.

(1) fail to reject
reject

Answers 0.0224

0.2576

(1) reject

D.

The coefficient is the difference between the population means of the milk production of cows with and without white spots.

ID: Concept Exercise 5.3.2

36. Using the textbook example of 420 California school districts and the regression of test scores on the student-teacher ratio, you find that the standard error on the slope coefficient is 0.51 when using the heteroskedasticity-robust formula, while it is 0.48 when employing the homoskedasticity-only formula.

When calculating the *t*-statistic, the recommended procedure is to:

- **A.** use the homoskedasticity-only formula because the *t*-statistic becomes larger.
- **B.** use the heteroskedasticity-robust formula.
- **C.** make a decision depending on how much different the estimate of the slope is under the two procedures.
- **D.** first test for homoskedasticity of the errors and then make a decision.

Answer: B. use the heteroskedasticity-robust formula.

ID: Test B Ex 5.4.5

37. Consider the following regression equation:

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

where X_i, Y_i, β_0, β_1, and μ_i denote the regressor, the regressand, the intercept coefficient, the slope coefficient, and the error term for the i^{th} observation, respectively.

When would the error term be homoskedastic?

- **A.** The error term is homoskedastic if the variance of the joint distribution of μ_i and Y_i is constant for $i = 1....n$, and in particular does not depend on Y_i.
- **B.** The error term is homoskedastic if the variance of the conditional distribution of μ_i given X_i is constant for $i = 1....n$, and in particular does not depend on X_i.
- **C.** The error term is homoskedastic if the variance of the conditional distribution of μ_i given Y_i is constant for $i = 1....n$, and in particular does not depend on Y_i.
- **D.** The error term is homoskedastic if the variance of the conditional distribution of μ_i given X_i is variable for $i = 1....n$, and in particular depends on X_i.

Which of the following statements describes the mathematical implications of heteroskedasticity?

- **A.** The OLS estimators remain unbiased, consistent, asymptotically normal, but do not necessarily have the least variance among all estimators that are linear in $Y_1, ..., Y_n$ conditional on $X_1, ..., X_n$.
- **B.** The OLS estimators remain unbiased, asymptotically normal and have the least variance among all estimators that are linear in $Y_1, ..., Y_n$, conditional on $X_1, ..., X_n$, but they are not consistent.
- **C.** The OLS estimators remain unbiased, consistent and have the least variance among all estimators that are linear in $Y_1, ..., Y_n$, conditional on $X_1, ..., X_n$, but they are not asymptotically normal.
- **D.** The OLS estimators remain consistent, asymptotically normal and have the least variance among all estimators that are linear in $Y_1, ..., Y_n$, conditional on $X_1, ..., X_n$, but they are not unbiased.

If the errors are heteroskedastic, then the *t*-statistic computed using (1) __________ standard error does not have a standard normal distribution, even in large samples.

(1)
- homoskedasticity-only
- heteroskedasticity-robust

Answers B.

The error term is homoskedastic if the variance of the conditional distribution of μ_i given X_i is constant for $i = 1....n$, and in particular does not depend on X_i.

A.

The OLS estimators remain unbiased, consistent, asymptotically normal, but do not necessarily have the least variance among all estimators that are linear in $Y_1, ..., Y_n$ conditional on $X_1, ..., X_n$.

(1) homoskedasticity-only

ID: Concept Exercise 5.4.1

38. Consider the following regression equation:

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

where X_i, Y_i, β_0, β_1, and μ_i denote the regressor, the regressand, the intercept coefficient, the slope coefficient, and the error term for the i^{th} observation, respectively.

Which of the following is the formula for calculating the heteroskedastic-robust estimator of the variance of $\hat{\beta}_1$?

A. $\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \dfrac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2 \hat{\mu}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)\right]^2}.$

B. $\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \dfrac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2 \hat{\mu}_i^2}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_i)\right]^2}.$

C. $\tilde{\sigma}^2_{\hat{\beta}_1} = \dfrac{\frac{1}{n-2}\sum_{i=1}^{n}\mu_i^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$

D. $\tilde{\sigma}^2_{\hat{\beta}_1} = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}\mu_i^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$

In practical econometric applications, it is better to assume that the errors might be (1) ____________ unless you have compelling reasons to believe otherwise.

Suppose that the errors in a regression model are heteroskedastic.

In large samples, the probability that a confidence interval constructed as ± 1.96 homoskedasticity-only standard errors contains the true value of the coefficient (2) ____________ 95%.

(1) heteroskedastic
homoskedastic

(2) will be
will not be

Answers

A. $\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \dfrac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2 \hat{\mu}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)\right]^2}.$

(1) heteroskedastic

(2) will not be

ID: Concept Exercise 5.4.2

39. If the errors are heteroskedastic, then:

- **A.** OLS is BLUE.
- **B.** LAD is BLUE if the conditional variance of the errors is known up to a constant factor of proportionality.
- **C.** OLS is efficient.
- **D.** WLS is BLUE if the conditional variance of the errors is known up to a constant factor of proportionality.

Answer: D. WLS is BLUE if the conditional variance of the errors is known up to a constant factor of proportionality.

ID: Test A Ex 5.5.5

40. Which of the following statements describe what the Gauss-Markov theorem states?

- ○ **A.** If the three least square assumptions hold and if errors are heteroskedastic, then the OLS estimator of a given population parameter will not be consistent, but it is still the most efficient linear unbiased estimator.
- ○ **B.** If the three least square assumptions hold and if errors are homoskedastic, then the OLS estimator of a given population parameter is the most efficient linear conditionally unbiased estimator.
- ○ **C.** If the three least square assumptions hold but if errors are not homoskedastic, then the OLS estimator of a given population parameter is not unbiased, but it is still the most efficient linear estimator.
- ○ **D.** If errors are homoskedastic but if the first least square assumption does not hold, then the OLS estimator of a given population parameter will not be linear, but it is still the most efficient unbiased estimator.

What are the limitations of the Gauss-Markov theorem? (*Check all that apply.*)

- ☐ **A.** The Gauss-Markov theorem does not provide a theoretical justification for using OLS.
- ☐ **B.** If the error term is heteroskedastic, then the OLS estimator may or may not be BLUE.
- ☐ **C.** Even if the conditions of the theorem hold, it is possible that under some conditions, other estimators that are not linear or conditionally unbiased may be more efficient than OLS.
- ☐ **D.** If the error term is heteroskedastic, then the OLS estimator is not BLUE.

In which of the following cases would the weighted least squares estimator (WLS) or the least absolute deviations estimator (LAD) be preferred to the OLS estimator? (*Check all that apply.*)

- ☐ **A.** LAD is preferred to OLS if the least square estimator is inconsistent.
- ☐ **B.** LAD is preferred to OLS if extreme outliers are not rare in the data.
- ☐ **C.** WLS is preffered to OLS if the errors are heteroskedastic.
- ☐ **D.** WLS is preffered to OLS if the least squares estimator is biased.

Answers B.

If the three least square assumptions hold and if errors are homoskedastic, then the OLS estimator of a given population parameter is the most efficient linear conditionally unbiased estimator.

C.

Even if the conditions of the theorem hold, it is possible that under some conditions, other estimators that are not linear or conditionally unbiased may be more efficient than OLS.
, D. If the error term is heteroskedastic, then the OLS estimator is not BLUE.

B. LAD is preferred to OLS if extreme outliers are not rare in the data., C.
WLS is preffered to OLS if the errors are heteroskedastic.

ID: Concept Exercise 5.5.1

---

41. Suppose that (Y_i, X_i) satisfy the assumptions specified here[13] and in addition, u_i is $N\left(0, \sigma_u^2\right)$ and independent of X_i. A random sample of n = 25 is drawn and yields

$$\underset{}{\hat{Y}} = \underset{(13.2)}{45.83} + \underset{(6.5)}{67.76}X,\ R^2 = 0.81,\ SER = 1.9$$

Where the numbers in parentheses are the homoskedastic-only standard errors for the regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Refer to the student t distribution with $n - 2$ degrees of freedom to answer the following questions.

Construct a 95% confidence interval for β_0 using the student t distribution (with $n - 2$ degrees of freedom) table available here[14].

The 95% confidence interval for β_0 is [____ , ____].

(*Round your responses to two decimal places*)

Suppose you wanted to test the hypothesis that β_1 equals 55 at the 5% level. That is,

$$H_0: \beta_1 = 55 \text{ vs. } H_1: \beta_1 \neq 55$$

Report the t-statistic for this test.

The t-statistic is ____ .

(*Round your response to two decimal places*)

Based on the t-statistic computed above, would you reject the null hypothesis at the 5% level?

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

Suppose you wanted to test that β_1 equals 55 versus the alternative that β_1 is greater than 55 at the 5% level. That is,

$$H_0: \beta_1 = 55 \text{ vs. } H_1: \beta_1 > 55$$

Based on the t-statistic of 1.96 computed above, would you reject the null hypothesis at the 5% level?

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

13: Definition

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: $E\left(u_i | X_i\right) = 0$;
2. $\left(Y_i, X_i\right)$, $i = 1, \ldots, n$, 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.

14: Definition

Critical Values for Student *t* Distribution

Degrees of freedom	One-sided 5%	Two-sided 5%
1	6.314	12.710
2	2.920	4.303
3	2.353	3.182
4	2.132	2.776
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.812	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145
15	1.753	2.131
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048
29	1.699	2.045
30	1.697	2.042

Answers 18.52

73.14

1.96

B. No.

A. Yes.

ID: Exercise 5.8

42. A professor wants to understand the relationship between students' class attendance and their academic performance. The professor selects 10 students at random for the experiment. Considering the least squares assumptions hold, the errors are homoskedastic, and the errors are normally distributed, the professor estimates the following regression equation:

$$\hat{Y}_i = 40 + 1.14X_i,$$

where Y_i measures the grade secured by the i^{th} student in the final exam and X_i measures the number of classes the student attends.

Which of the following statements is true for the components of the homoskedasticity-only t-statistic testing the hypothesis $H_0: \beta_1 = \beta_{1,0}$ vs. $H_1: \beta_1 \neq \beta_{1,0}$?

- A. $\left(\hat{\beta}_1 - \beta_{1,0}\right)$ and $\tilde{\sigma}^2_{\hat{\beta}_1}$ both follow a chi-squared distribution with $n-1$ degrees of freedom.
- B. $\left(\hat{\beta}_1 - \beta_{1,0}\right)$ follows a chi-squared distribution with $n-1$ degrees of freedom while $\tilde{\sigma}^2_{\hat{\beta}_1}$ follows a standard normal distribution.
- C. $\left(\hat{\beta}_1 - \beta_{1,0}\right)$ follows a standard normal distribution while $\tilde{\sigma}^2_{\hat{\beta}_1}$ follows a chi-squared distribution with $n-2$ degrees of freedom.
- D. $\left(\hat{\beta}_1 - \beta_{1,0}\right)$ and $\tilde{\sigma}^2_{\hat{\beta}_1}$ both follow a standard normal distribution.

Suppose the homoskedasticity-only standard error of the OLS estimator of the slope is 0.55.

The homoskedasticity-only t-statistic associated with the test $H_0: \beta_1 = 0.08$ vs. $H_1: \beta_1 \neq 0.08$ is ____.

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

Based on the t-statistic, we (1) ____ the null hypothesis at the 5% significance level.

(1) ○ reject
○ fail to reject

Answers C.

$\left(\hat{\beta}_1 - \beta_{1,0}\right)$ follows a standard normal distribution while $\tilde{\sigma}^2_{\hat{\beta}_1}$ follows a chi-squared distribution with $n-2$ degrees of freedom.

1.93

(1) fail to reject

ID: Concept Exercise 5.6.1