Name: Answers

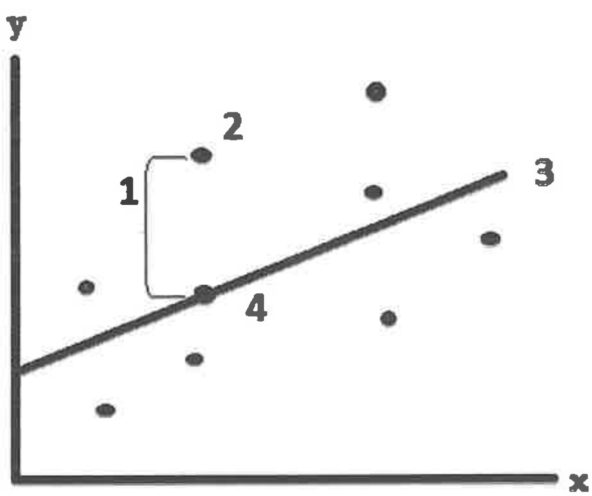
EC380: Econometrics Exam 1

Scott Hiller

Spring 2022

For multiple choice questions, choose the answer that *best* answers the question. For calculation problems, show all of your work and be sure to identify any assumptions you are making. Good luck!

1. Consider the **sample** regression function depicted below:



Match the letter options to the number labels.

1. A
2. F
3. H
4. E

A. $\hat{u}_i$
B. u_i
C. x_i
D. $\hat{x}_i$
E. $\hat{y}_i$
F. y_i
G. $\beta_0 + \beta_1 x_i$
H. $\hat{\beta}_0 + \hat{\beta}_1 x_i$

2. When we use Ordinary Least Squares to estimate a model such as $y_i = \beta_0 + \beta_1 x_i + u_i$, we are estimating

(a) each value y_i

(b) each value x_i

(c) the mean of y conditional on x

(d) the mean of x conditional on y

3. Consider the regression model

$$GPA_i = \beta_0 + \beta_1 Log(HoursStudied_i) + u_i$$

Which of the following is a correct interpretaton?

(a) 1% more of hours studied yields on average $\beta_1/100$ point increase in GPA.
(b) 1 more hour studied yields on average β_1 additional points of GPA.
(c) β_1 more hours studied yields on average 1 more point of GPA.
(d) 1 more hour studied yields on average $100 * \beta_1$ percent increased GPA.

4. Suppose you estimate the regression model using data on 50 different countries:

$$GDP_i = \beta_0 + \beta_1 Literacy_i + \beta_2 Democracy_i + u_i$$

Your Stata output includes $R^2 = 0.6$, which tells you that

(a) this model is explaining 60% of the variation in GDP.
(b) the explained sum of squares divided by the total sum of squares is 40%.
(c) the residual sum of squares divided by the total sum of squares is 60%.
(d) literacy explains 60% of the variation in GDP.

5. Which of the following is NOT an appropriate model for OLS estimation?

(a) $log(y_i) = \beta_0 + \beta_1 x_i + u_i$
(b) $y_i = \beta_0 + \beta_1 \sqrt{x_i} + u_i$
(c) $y_i = \beta_0 + \beta_1 x_i^3 + u_i$
(d) $y_i = \frac{1}{\beta_0} + \sqrt{\beta_1} x_i + u_i$

Consider the model for the next two questions

$$Salary_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + \beta_3 (Education_i * Experience_i) + u_i$$

which includes an interaction term.

6. You are including the <u>interaction</u> of education and experience in this regression model because you think that

(a) education and experience are correlated.
(b) education and experience each has some effect on salary.
(c) education and experience have some overlapping effect on salary.
(d) there are other variables relevant to explaining salary which have been omitted from the model.

7. Suppose Stata gives you the following estimates:
$\beta_0 = 215, \beta_1 = 220, \beta_2 = 540, \beta_3 = 25$
For someone with 14 years of experience, we expect 1 additional year of education to yield

   (a) $[215 + 220 - (540 * 10) + (25 * 12)]$ additional dollars of salary.

   (b) $[220 + 25]$ additional dollars of salary.

   (c) $[(540 * 10) + (25 * 14)]$ additional dollars of salary.

   (d) $[220 + (25 * 14)]$ additional dollars of salary.

8. Consider the model: $Salary_i = \beta_0 + \beta_1 Education_i + \beta_2 Age_i + \beta_3 Age^2 + u_i$ where salary is measured in dollars and age and education are measured in years. The effect of being 1 year older on salary is

   (a) $\beta_3 Age$

   (b) β_2

   (c) $\beta_2 + \beta_3$

   (d) $\beta_2 + 2\beta_3 Age$

9. Consider the model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$. The variance of $\hat{\beta}_1$ gives us a measure of the precision of $\hat{\beta}_1$ as an estimator of β_1. $\hat{\beta}_1$ is a less precise estimator of β_1 when

   (a) the sample size is larger.

   (b) the variance of the errors is small.

   (c) there is more variation in the x values.

   (d) x_1 and x_2 are more correlated.

10. OLS *fails* to yield an unbiased estimate of β_1 if which of the following is true?

   (a) There is a perfect linear relationship between two of the explanatory variables.

   (b) There is a large amount of variation in the values of each explanatory variable.

   (c) There are at least as many observations as parameters β being estimated.

   (d) The expected value of the errors conditional on the explanatory variables is NOT zero, i.e. $E(u|x) \neq 0$.

11. Our OLS estimators are ____ if $var(\hat{\beta})$ approaches zero as the sample size gets very large $(n \to \infty)$.

   (a) linear

   (b) consistent B

(c) efficient

(d) unbiased

12. Suppose you have a dataset containing income and consumption expenditures for 10,000 individuals. After exploring the data you observe that some high-income individuals consume a lot while other high-income individuals consume a little, but low-income individuals are constrained to low levels of consumption. Which assumption of the classical linear regression model is violated when you regress $Consumption_i = \beta_0 + \beta_1 Income_i + u_i$?

(a) The errors have a mean of zero, $E(u_i) = 0$.

(b) There is no perfect multicollinearity among the explanatory variables.

(c) The errors are homoskedastic, $var(u_i) = \sigma^2$.

(d) There is zero correlation across observations, $cov(u_i, u_j) = 0$.

13. The Gauss-Markov Theorem tells us that under the assumptions of the classical linear model, the OLS estimators are BLUE, or Best Linear Unbiased Estimators. "Best" means

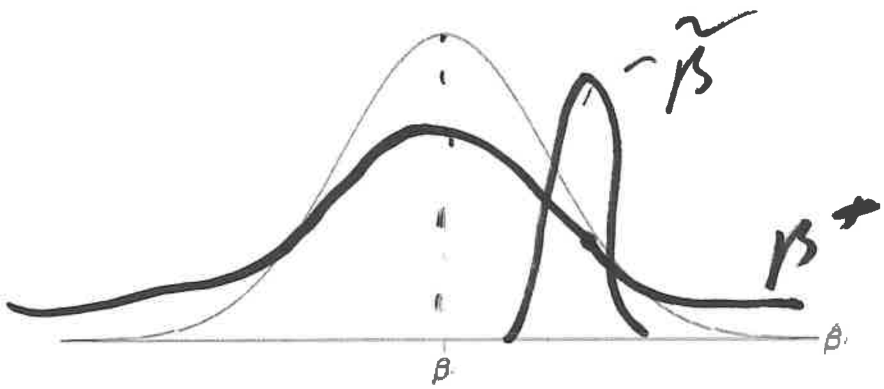
(a) on average OLS will estimate the true population β correctly.

(b) $\hat{\beta}$, the OLS estimate of β, is more precise than any other unbiased linear estimators.

(c) the OLS estimate $\hat{\beta}$ equals β, the true population parameter.

(d) $\hat{\beta}$, the OLS estimate of β, has the largest variance of any unbiased linear estimators.

14. Consider the following distribution of 1,000 $\hat{\beta}$s that were estimated from 1,000 samples randomly drawn from the population. On the existing figure, draw the distribution of another estimator, $\beta*$, that is unbiased but "less efficient" than the OLS estimator $\hat{\beta}$. Also, draw a distribution, $\tilde{\beta}$ that is more efficient, but biased. Label clearly.



15. **Hypothesis Testing** Suppose you run the following regression relating economic growth of a country (Growth) to how educated its population is (Educ), how healthy its population is as captured by child mortality rates (Mort), and its trade volume (Trade).

$$Growth_i = \beta_0 + \beta_1 Educ_i + \beta_2 Mort_i + \beta_3 Trade_i + u_i$$

Your regression results are:

$$\begin{aligned} \hat{Growth}_i &= 2.73 + 2.14 Educ_i - 1.76 Mort_i + 0.83 Trade_i \\ se &= (0.26) \quad (0.76) \qquad (0.54) \qquad (0.47) \\ R^2 &= 0.42 \\ RSS &= 0.23 \\ n &= 44 \text{countries} \end{aligned}$$

(a) Test whether $\hat{\beta}_2$ is statistically significant using the **confidence interval approach.** Carefully state your hypothesis. Explicitly state your conclusions. Make sure to specify what went into the determination of the critical value.

$H_0: \hat{\beta}_2 = 0$

$H_a: \hat{\beta}_2 \neq 0$

$\hat{\beta}_2 \pm se(\hat{\beta}_2)\, t_{\alpha/2}$

$\hat{\beta}_2 - se(\hat{\beta}_2) < \beta_2 < \hat{\beta}_2 + se(\hat{\beta}_2)$

$-1.76 - .54(2) < \beta_2 < -1.76 + .54(2)$

$-.68 \geq \beta_2 \geq -2.14$

reject H_0

(b) Use the t-test approach to test whether the parameter on *Trade* is negative. Show all your work.

$H_0: \hat{\beta}_3 < 0$

$H_a: \hat{\beta}_3 > 0$

$\frac{.8370}{.47} = 1.766$

Reject H_0

16. Consider the model $y_i = \beta_0 + \beta_1 x_i + u_i$. Use either the "method of moments" (using assumptions) or the "method of least squares" to derive the OLS estimator for β_1, the slope parameter. Show all of your work. (Hint: $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$.)

I ∩ ∩ [illegible]

17. Prove that the OLS estimator $\hat{\beta}_1$ you derived above is an unbiased estimator of β_1. Show all of your work and state any assumptions being made.

Copy from notes.

18. Consider the following regression model:

$$Price_{Tequila} = \beta_0 + \beta_1 Price_{Agave} + \beta_2 Price_{Limes} + \beta_3 Price_{Rum} + u_i$$

where agave is the main ingredient in tequila, limes and tequila are complements (consumed together), and rum is a substitute for tequila.

(i) Income is a relevant variable that you have left out of the model. If income is positively correlated with the price of tequila and negatively correlated with the price of rum, we know that our estimate of β_3 is

(a) unbiased

(b) biased downward

(c) biased upward

(d) biased but we don't know in which direction (up or down)

(ii) Suppose you add income to the model and you observe that the variance of $\hat{\beta}_3$ rises. Explain why $var(\hat{\beta}_3)$ rose.

$$var(\hat{\beta}_3) = \frac{var(u_i)}{\sum (x_3 - \bar{x}_3)(1 - r_3^2)}$$

r^2 increases b/c income

+ price of rum are correlated.

Now less variation in x_3. $\hat{\beta}_3$ less precise.

Bonus! Explain the main differences between these two econometric problems: serial correlation and heteroskedasticity.