Econometrics
Review questions for exam

Nathaniel Higgins
nhiggins@jhu.edu,

1.

Suppose you have a model:

$$y = \beta_0 x_1 + u$$

You propose the model above and then estimate the model using OLS to obtain:

$$\widehat{y} = \widehat{\beta}_0 x_1.$$

Label each of the items in the preceding two equations as either random, potentially random, or definitely deterministic.

2.

Your computer is broken. You need to determine the OLS estimator for a simple model:

$$y = \beta_0 + \beta_1 x_1 + u.$$

Luckily, the data set is small. It is displayed below. To the nearest whole digit (no need for decimals) find the OLS estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

y	x
2	1
5	2
6	3
7	3
3	1

3.

Type II error refers to the following case:
(a) reject the null when the null is false
(b) reject the null when the null is true
(c) fail to reject the null when the null is false
(d) fail to reject the null when the null is true

4.

Which of the following statements about the OLS residuals is not true?
(a) The sample covariance between the residuals and the regressors equals zero.
(b) The sum of the residuals equals zero.
(c) The sample covariance between the residuals and the dependent variable equals zero.
(d) The sample average of the residuals equals zero.

5.

Which expression correctly describes the predicted value ($\widehat{y}$) of a multivariate regression model?
(a) $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$
(b) $\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{u}$
(c) $\beta_0 + \beta_1 x_1 + \beta_2 x_2$
(d) $\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$

6.

True or false: In the multiple regression model, the goodness-of-fit measure R-squared always increases or remains the same when an additional explanatory variable is added.

7.

True or false: If the error term in a linear regression model is normally distributed, then the OLS estimator is the best linear unbiased estimator.

8.

True or false and explain: Suppose you are interested in estimating the relationship between birth weight and prenatal visits. You have data on birth weight, in pounds, number of prenatal visits, and the age and race of the mother. However, you do not observe family income. As a result, the coefficient on prenatal visits will have a positive bias.

9.

True or false and explain: If there is a lot of variation in the data for a regressor, then the variance of the coefficient on that regressor will be smaller than if there is not a lot of variation in the data for that regressor.

10.

Suppose we are interested in estimating a simple model that relates the number of crimes on college campuses (crime) to student enrollment (enroll).
(a) Set up a constant elasticity (log-log) model that explains the number of crimes on a college campus using student enrollment.
(b) Suppose that after running the model from part (a) on a random sample of 500 colleges, you find a coefficient of 1.27 (standard error of 0.11) on the variable associated with enrollment. Interpret this coefficient.
(c) Next you would like to know whether crime is more of a problem on larger campuses than smaller campuses. Write down a hypothesis test that examines whether a one percent increase in student enrollment is associated with an increase of more than one percent in the number of crimes.
(d) Larger schools may be located in areas with higher crime rates, e.g., urban centers. Discuss the potential of omitted variable bias in the model from part (a). Derive the sign of the bias; be explicit about your assumptions.

11.

Suppose that a regression of college GPA (in points) on high school GPA (in points), a dummy for male, and a dummy for computer ownership yields the following 95% confidence interval for the estimated regression slope coefficient on high school GPA (reported by Stata): $\widehat{\beta_1} \in (0.292, 0.667)$. What would the result be of a test of the null hypothesis that $\beta_1 = 0$ against a two-sided alternative.

12.

Two explanatory variables are included in a multivariate model. They are positively correlated. You run an OLS regression model and obtain:

$$\widehat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2$$

You then run a regression of the same dependent variable on each of the independent variables individually. Call the coefficients on x_1 and x_2 in these individual regressions δ_1 and δ_2, respectively. The mean and std. dev. of each of the variables are shown in the table below

variable	mean	std. dev.
x1	5	2
x2	3	6

Which of the coefficients changes more? I.e., which is bigger in absolute value: $\widehat{\beta_1} - \widehat{\delta_1}$ or $\widehat{\beta_2} - \widehat{\delta_2}$?

13.

Suppose your R-squared is 0.30, and the explained sum of squares is 16. What is the unexplained sum of squares?

14.

Consider two estimated slope coefficients ($\widehat{\beta_1}$ and $\widehat{\delta_1}$) that result from the following regression models:

$$y = \beta_0 + \beta_1 x_1 + u$$
$$y = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + v$$

Explain the circumstances under which $\widehat{\beta_1} = \widehat{\delta_1}$.

15.

$$E(GPA|SAT) = 0.70 + 0.002SAT$$

The mean and std. dev. of GPA and SAT are:

variable	mean	std. dev.
GPA	2.9	0.45
SAT	1050	300

What is the unconditional expectation of GPA?

16.

Consider two estimators of y: z_1 and z_2. z_1 is unbiased. z_2 is biased. The variance of z_1 is 3. The variance of z_2 is 1. Which estimator do you prefer? Why?

17.

When building models like:

$$y = \beta_0 + \beta_1 x_1 + u,$$

we usually assume that $E(u) = 0$. When we estimate the model above by OLS, is it important that $E(u) = 0$? What about $E(u|x_1)$? Can $E(u|x_1)$ be equal to something other than 0?

18.

What is multicollinearity? Why is it scary? Why is it not that scary?

19.

If Stata reports that the p-value associated with a variable in an OLS regression is 0.06, would you reject the null hypothesis that the coefficient associated with this variable is equal to 0 at the 10% level?

20.

Recall that the variance of the OLS estimator is given by:

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}.$$

Suppose the std. dev. of x_j, an independent variable in an OLS regression model, was 3.45. You then went and got more data (you increased N). After adding more observations, the std. dev. of x_j in your sample was 3.0. What happens to the variance of $\widehat{\beta}_j$?

21.

Suppose you estimated a model by OLS and obtained the equation:

$$\widehat{y} = 10 + 4x_1 - 6x_2.$$

What happens to $\widehat{\beta}_1$ if you multiply x_1 by 10?

22.

Suppose that you have estimated a model by OLS and you obtain residuals (predictions of the unobservable component of the model, usually denoted by u). The table of residuals is

-0.6
0.4
-0.6
0.4
–

What is the value of the missing residual?