

A Sample Mid-Term Examination of Econometrics

Multiple Choice

1) The conditional expectation of Y given X , $E(Y | X = x)$, is calculated as follows:

A) $\sum_{i=1}^k Y_i \Pr(X = x_i | Y = y)$

B) $E[E(Y | X)]$

C) $\sum_{i=1}^k y_i \Pr(Y = y_i | X = x)$

D) $\sum_{i=1}^k E(Y | X = x_i) \Pr(X = x_i)$

2) Assume that Y is normally distributed $N(\mu, \sigma^2)$. Moving from the mean (μ) 1.96 standard deviations to the left and 1.96 standard deviations to the right, then the area under the normal p.d.f. is

A) 0.67

B) 0.05

C) 0.95

D) 0.33

3) An estimator $\hat{\mu}_Y$ of the population value μ_Y is unbiased if

A) $\hat{\mu}_Y = \mu_Y$.

B) \bar{Y} has the smallest variance of all estimators.

C) $\bar{Y} \xrightarrow{p} \mu_Y$.

D) $E(\hat{\mu}_Y) = \mu_Y$.

4) Among all unbiased estimators that are weighted averages of Y_1, \dots, Y_n , \bar{Y} , is

A) the only consistent estimator of μ_Y .

B) the most efficient estimator of μ_Y .

C) a number which, by definition, cannot have a variance.

D) the most unbiased estimator of μ_Y .

5) Assume that you have 125 observations on the height (H) and weight (W) of your peers in college.

Let $s_{HW} = 68$, $s_H = 3.5$, $s_W = 29$. The sample correlation coefficient is

- A) 1.22
- B) 0.50
- C) 0.67
- D) Cannot be computed since males and females have not been separated out.

6) When the estimated slope coefficient in the simple regression model, $\hat{\beta}_1$, is zero, then

- A) $R^2 = \bar{Y}$.
- B) $0 < R^2 < 1$.
- C) $R^2 = 0$.
- D) $R^2 > (SSR/TSS)$.

7) The reason why estimators have a sampling distribution is that

- A) economics is not a precise science.
- B) individuals respond differently to incentives.
- C) in real life you typically get to sample many times.
- D) the values of the explanatory variable and the error term differ across samples.

8) The variance of Y_i is given by

A) $\beta_0^2 + \beta_1^2 \text{ var}(X_i) + \text{ var}(u_i)$.

B) the variance of u_i .

C) $\beta_1^2 \text{ var}(X_i) + \text{ var}(u_i)$.

D) the variance of the residuals.

9) To obtain the slope estimator using the least squares principle, you divide the

- A) sample variance of X by the sample variance of Y .
- B) sample covariance of X and Y by the sample variance of Y .
- C) sample covariance of X and Y by the sample variance of X .
- D) sample variance of X by the sample covariance of X and Y .

10) Changing the units of measurement, e.g. measuring test scores in 100s, will do all of the following EXCEPT for changing the

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

11) Heteroskedasticity means that

- A) homogeneity cannot be assumed automatically for the model.
- B) the variance of the error term is not constant.
- C) the observed units have different preferences.
- D) agents are not all rational.

12) The t -statistic is calculated by dividing

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

13) The 95% confidence interval for β_1 in large samples is the interval

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

14) Consider the estimated equation from your textbook

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, R^2 = 0.051, SER = 18.6$$

(10.4) (0.52)

The t -statistic for the slope is approximately

- A) 4.38
- B) 67.20
- C) 0.52
- D) 1.76

15) 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 possible decisions is the only correct one:

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

16) Under imperfect multicollinearity

- A) the OLS estimator cannot be computed.

- B) two or more of the regressors are highly correlated.
- C) the OLS estimator is biased even in samples of $n > 100$.
- D) the error terms are highly, but not perfectly, correlated.

17) When you have an omitted variable problem, the assumption that $E(u_i | X_i) = 0$ is violated. This implies that

- A) the sum of the residuals is no longer zero.
- B) there is another estimator called weighted least squares, which is BLUE.
- C) the sum of the residuals times any of the explanatory variables is no longer zero.
- D) the OLS estimator is no longer consistent.

18) In the multiple regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$, $i = 1, \dots, n$, the OLS estimators are obtained by minimizing the sum of

- A) squared mistakes in $\sum_{i=1}^n (Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki})^2$
- B) squared mistakes in $\sum_{i=1}^n (Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki} - u_i)^2$
- C) absolute mistakes in $\sum_{i=1}^n |(Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki})|$
- D) squared mistakes in $\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2$

19) In the multiple regression model, the SER is given by

- A) $\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i$
- B) $\frac{1}{n-k-2} \sum_{i=1}^n u_i$
- C) $\frac{1}{n-k-2} \sum_{i=1}^n \hat{u}_i$
- D) $\frac{1}{n-k-2} \sum_{i=1}^n \hat{u}_i^2$

20) Consider the following multiple regression models (a) to (d) below. $DFemme = 1$ if the individual is a female, and is zero otherwise; $DMale$ is a binary variable which takes on the value one if the individual is male, and is zero otherwise; $DMarried$ is a binary variable which is unity for married individuals and is zero otherwise, and $DSingle$ is $(1-DMarried)$.

Regressing weekly earnings ($Earn$) on a set of explanatory variables, you will experience perfect multicollinearity in the following cases unless:

- A) $\widehat{Eam}_i = \widehat{\beta}_0 + \widehat{\beta}_1 DFemme + \widehat{\beta}_2 Dmale + \widehat{\beta}_3 X_{3i}$
- B) $\widehat{Eam}_i = \widehat{\beta}_0 + \widehat{\beta}_1 DMarried + \widehat{\beta}_2 DSingle + \widehat{\beta}_3 X_{3i}$
- C) $\widehat{Eam}_i = \widehat{\beta}_0 + \widehat{\beta}_1 DFemme + \widehat{\beta}_3 X_{3i}$

$$D) \widehat{Eam}_i = \widehat{\beta}_1 DFemme + \widehat{\beta}_2 Dmale + \widehat{\beta}_3 DMarried + \widehat{\beta}_4 DSingle + \widehat{\beta}_5 X_{3i}$$

Long Questions

1. The OLS formula for the slope coefficients in the multiple regression model become increasingly more complicated, using the "sums" expressions, as you add more regressors. For example, in the regression with a single explanatory variable, the formula is

$$\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Show that for a multiple regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

the formula for the slope of the first explanatory variable is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_{1i} \sum_{i=1}^n x_{2i}^2 - \sum_{i=1}^n y_i x_{2i} \sum_{i=1}^n x_{1i} x_{2i}}{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2 - \left(\sum_{i=1}^n x_{1i} x_{2i} \right)^2}$$

(small letters refer to deviations from means as in $z_i = Z_i - \bar{Z}$. Hint: use the FWL theorem to reduce the model to a multiple regression model without an intercept. Then use the matrix formula of $\hat{\beta}$ for the reduced model.)

2 Sir Francis Galton, a cousin of James Darwin, examined the relationship between the height of children and their parents towards the end of the 19th century. It is from this study that the name "regression" originated. You decide to update his findings by collecting data from 110 college students, and estimate the following relationship:

$$\widehat{Studenth} = 19.6 + 0.73 \times Midparh, R^2 = 0.45, SER = 2.0$$

(7.2) (0.10)

where *Studenth* is the height of students in inches, and *Midparh* is the average of the parental heights. Values in parentheses are heteroskedasticity robust standard errors. (Following Galton's methodology, both variables were adjusted so that the average female height was equal to the average male height.)

(a) Test for the statistical significance of the slope coefficient.

- (b) If children, on average, were expected to be of the same height as their parents, then this would imply two hypotheses, one for the slope and one for the intercept.
- (i) What should the null hypothesis be for the intercept? Calculate the relevant t -statistic and carry out the hypothesis test at the 1% level.
- (ii) What should the null hypothesis be for the slope? Calculate the relevant t -statistic and carry out the hypothesis test at the 5% level.
- (c) Can you reject the null hypothesis that the regression R^2 is zero?
- (d) Construct a 95% confidence interval for a one inch increase in the average of parental height.

Answers

Multiple Choice

- 1) C
- 2) C
- 3) D
- 4) B
- 5) C
- 6) C
- 7) D
- 8) C
- 9) C
- 10) C
- 11) B
- 12) C
- 13) C
- 14) A
- 15) B
- 16) B
- 17) D
- 18) A
- 19) D
- 20) C

Long Questions

1. We will demonstrate this question in class.

2.

(a) $H_0 : \beta_1 = 0$, $t=7.30$, for $H_1 : \beta_1 > 0$, the critical value for a two-sided alternative is 1.645.

Hence we reject the null hypothesis

(b) $H_0 : \beta_0 = 0$, $t=2.72$, for $H_1 : \beta_0 \neq 0$, the critical value for a two-sided alternative is 2.58.

Hence we reject the null hypothesis in (i). For the slope we have $H_0 : \beta_1 = 1$, $t=-2.70$, for $H_1 : \beta_1 \neq 1$, the critical value for a two-sided alternative is 1.96. Hence we reject the null hypothesis in (ii).

(c) For the simple linear regression model, $H_0 : \beta_1 = 0$ implies that $R^2 = 0$. Hence it is the same test as in (a).

(d) $(0.73 - 1.96 \times 0.10, 0.73 + 1.96 \times 0.10) = (0.53, 0.93)$.