ECN 102, Spring 2020

Midterm 2 Review
Multiple Choice

Multiple Choice Question 1

The OLS estimator

(a) minimizes $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

(b) minimizes $\sum_{i=1}^{n}(y_i - \bar{y})^2$

(c) minimizes $\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$

(d) none of the above

Answer: a. OLS minimizes the sum of squared residuals (RSS). Remember, the residuals measure how far off the regression line is from the actual data, and we want a line that is as close a possible to the data. Minimizing how far off something is is equivalent to maximizing how close something is.

Multiple Choice Question 2

For linear regression, the conditional mean of y given $x = x^*$ equals

(a) $b_1 + b_2 x^*$

(b) $b_1 + b_2 x^* + e$

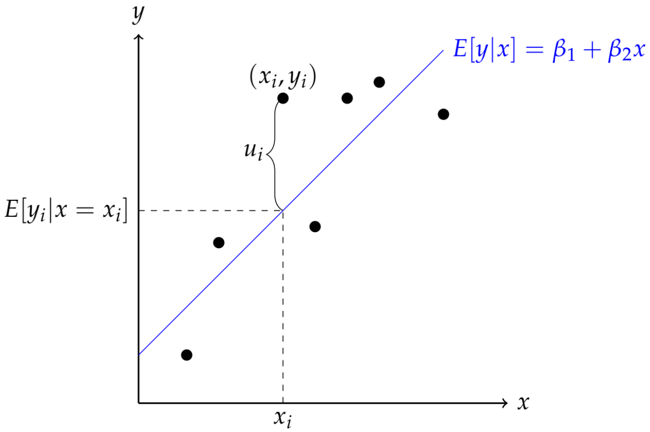
(c) $\beta_1 + \beta_2 x^*$

(d) $\beta_1 + \beta_2 x^* + u$

(e) none of the above

Answer: c. A line of best fit for the population tells us what we expect y to be for any value of x, expressed $E[y|x] = \beta_1 + \beta_2 x$. (The sample equivalent is called the **fitted value**, expressed $\hat{y} = b_1 + b_2 x$.) An illustration is on the next slide.

Multiple Choice Question 2



Multiple Choice Question 3

The standard error of the regression is a measure of

(a) the standard deviation of the slope coefficient

(b) the standard deviation of the intercept coefficient

(c) the standard deviation of the dependent variable

(d) the standard deviation of the error

(e) none of the above

Answer: d. The standard error of the regression is sometimes called the **standard error of the residual** or the **root mean square error (RMSE)**. That is because it's given by

$$s_e \equiv \sqrt{\frac{\text{RSS}}{n-2}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

which hopefully you recognize is really just a standard deviation formula.

Multiple Choice Question 4

We regress y on x and find that $b_2 = 10$ with standard error 2. Given only this information,

(a) the regressor x is highly statistically significant and highly economically significant

(b) the regressor x is highly statistically significant

(c) the regressor x is highly economically significant

(d) none of the above

Answer: b. *Statistically significant*: slope coefficient is statistically non-zero, i.e. we reject the null hypothesis $H_0 : \beta_2 = 0$. Here, $t = (10 - 0)/2 = 5$, which is rejected at conventional levels.

Economically significant: slope coefficient is non-zero enough that it has practical importance. Suppose this regression says that an extra year of schooling is associated with an extra 10 cents of annual income. Economically insignificant, even though it's statistically significant.

Multiple Choice Question 5

The standard error of the slope coefficient

(a) increases with increases in the sample size

(b) decreases with increases in the variability of the regressors

(c) both (a) and (b)

(d) neither (a) nor (b)

Answer: b. The standard error of the slope coefficient is given by

$$\mathrm{se}(b_2) = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

First, because OLS estimates are consistent, s_e gets smaller as sample size increases, so option (a) cannot be right.

Second, denominator is essentially standard deviation of the regressor. When there's more variation in the regressor, the denominator gets bigger, and therefore $\mathrm{se}(b_2)$ gets smaller.

Multiple Choice Question 6

The Stata command `regress y x, vce(robust)`

(a) yields the same *t*-statistics as command `regress y x`

(b) yields the same *p*-value as command `regress y x`

(c) both (a) and (b)

(d) neither (a) nor (b)

Answer: d. We use option `vce(robust)` when we suspect heteroskedasticity (i.e. most of the time in practice). This gives a different standard error, and therefore a different *t*-statistic (which is a function of the standard error), and therefore a different *p*-value (which is a function of the *t*-statistic).

Multiple Choice Question 7

If sample covariance is positive, then

(a) the sample correlation coefficient is necessarily positive

(b) the sample correlation coefficient is most likely positive, but could be negative

(c) the sample correlation coefficient could easily be positive or negative

Answer: a. The correlation coefficient is defined to be

$$r_{xy} = \frac{s_{xy}}{s_x s_y}.$$

Standard deviations cannot be negative, so if we divide by both positive numbers, then the sign of the entire fraction is still positive.

Multiple Choice Question 8

Regression of y on x yields slope coefficient 0.50 and correlation coefficient 0.40. It follows that regression of x on y using the same data yields

(a) slope coefficient 2.0

(b) correlation coefficient 0.40

(c) both (a) and (b)

(d) neither (a) nor (b)

Answer: b. The slope of the reverse regression is *not* the reciprocal of the slope we get from regression y on x. To see this, note

$$\texttt{regress y x} \quad \Longrightarrow \text{slope coefficient: } \frac{s_{xy}}{s_x^2}$$

$$\texttt{regress x y} \quad \Longrightarrow \text{slope coefficient: } \frac{s_{yx}}{s_y^2}$$

A bit unintuitive, this one. We know $s_{xy} = s_{yx}$ (and $r_{xy} = r_{yx}$), but there's no reason to assume any particular relationship between the variance of x and y in general, so there's no reason to conclude that s_{xy}/s_x^2 is the reciprocal of s_{xy}/s_y^2.

Multiple Choice Question 9

For $(b_2 - \beta_2)/\,\mathrm{se}(b_2)$ to be exactly $T(n-2)$ distributed, it is necessary that

(a) assumptions 1-4 hold

(b) assumptions 1-4 hold and the error term is normally distributed

(c) assumptions 1-4 hold and the error term is $T(n-2)$ distributed

Answer: b. Recapping the implications of the assumptions (OLS5 is normally distributed errors):

- OLS1-2: estimates are unbiased
- OLS1-4: estimates are consistent and BLUE
- OLS1-5: OLS estimates are BUE and $(b_2 - \beta_2)/\,\mathrm{se}(b_2)$ is drawn from exact $T(n-2)$ distribution

Multiple Choice Question 10

Suppose $b_2 = -5$ and $R^2 = 0.25$. Then the correlation coefficient is

(a) $r_{xy} = 0.5$

(b) $r_{xy} = 0.0625$

(c) not enough information

(d) none of the above

Answer: d. Recall that $R^2 = r_{xy}^2$. Therefore

$$r_{xy}^2 = 0.25 \quad \Longrightarrow \quad r_{xy} = \sqrt{0.25} = \pm 0.5.$$

Okay, there's a plus and a minus square root. Because the slope coefficient is negative, it must be the case that $r_{xy} = -0.5$.