ECON471
Fall 2020
Problem Set 2
Due Wednesday September 30, by 11:59pm CST

Name: Wenxiao Yang

Section: B3

1. Show that the R^2 from the regression of Y on X is the same as the R^2 from the regression of X on Y.

$$y = \beta_0 + \beta_1 x + u \qquad x = \alpha_0 + \alpha_1 y + e$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \qquad \bar{x} = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{y}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 y_i$$

$$\hat{y}_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x}) \qquad \hat{x}_i - \bar{x} = \hat{\alpha}_1 (y_i - \bar{y})$$

$$\hat{\alpha}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} = \hat{\beta}_1 \cdot \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}$$

$$\Rightarrow \frac{\hat{\alpha}_1}{\hat{\beta}_1} \frac{\sum (y_i - \bar{y})^2}{\sum (x_i - \bar{x})^2} = 1$$

$$R^2_{Y \text{ on } X} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\hat{\alpha}_1^2 \sum (\hat{y}_i - \bar{y})^2}{\sum (\hat{x}_i - \bar{x})^2}$$

$$R^2_{X \text{ on } Y} = \frac{\sum (\hat{x}_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \frac{\hat{\beta}_1^2 \sum (\hat{x}_i - \bar{x})^2}{\sum (\hat{y}_i - \bar{y})^2}$$

$$\Rightarrow \frac{R^2_{Y \text{ on } X}}{R^2_{X \text{ on } Y}} = \left(\frac{\hat{\alpha}_1}{\hat{\beta}_1} \cdot \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (\hat{x}_i - \bar{x})^2} \right)^2 = 1.$$

$$\Rightarrow R^2_{Y \text{ on } X} = R^2_{X \text{ on } Y}$$

2. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + u_i$. Given the n observations $\{(x_i, y_i),\ i = 1,2, \dots\dots, n\}$, we estimate β_1 as

$$\tilde{\beta}_1 = \frac{1}{n-1}\sum_{i=2}^{n}\frac{(y_i - y_{i-1})}{(x_i - x_{i-1})}$$

a. Give a geometric interpretation of $\tilde{\beta}_1$.

when x_{i-1} change into next x_i, $\Delta x = x_i - x_{i-1}$, the y_{i-1} will change to y_i, $\Delta y = y_i - y_{i-1}$, $\frac{\Delta y}{\Delta x} = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$ is the slope in x_{i-1} to x_i

So the $\tilde{\beta}_1$ is the averge slope for x_1 to x_2, --- x_{n-1} to x_n

b. Show that $\tilde{\beta}_1$ is an unbiased estimator of β_1. Be sure to state the assumptions needed to prove this.

$$E(\tilde{\beta}_1) = \frac{1}{n-1}\sum_{i=2}^{n} E\,\frac{(y_i - y_{i-1})}{(x_i - x_{i-1})}$$

$$= \frac{1}{n-1}\sum_{i=2}^{n} E\,\frac{(\beta_0 + \beta_1 x_i + u_i) - (\beta_0 + \beta_1 x_{i-1} + u_{i-1})}{x_i - x_{i-1}}$$

$$= \frac{1}{n-1}\sum_{i=2}^{n} E\left(\beta_1 + \frac{u_i - u_{i-1}}{x_i - x_{i-1}}\right)$$

$$= \frac{1}{n-1}(n-1)\beta_1 + \frac{1}{n-1}\sum_{i=2}^{n}\frac{E(u_i - u_{i-1})}{x_i - x_{i-1}}$$

(Since $E(u_i|x_i) = 0$, we know $E(u_i - u_{i-1}|x_i, x_{i-1}) = 0$)

(Since $\{(x_i, y_i) | i = 1, \dots, n\}$ is random sample, we know $E(u_i - u_{i-1}) = E(E(u_i - u_{i-1}|x_i, x_{i-1})) = 0$.)

$$\Rightarrow E(\tilde{\beta}_1) = \beta_1 + 0 = \beta_1$$

$\Rightarrow \tilde{\beta}_1$ is an unbiased estimator of β_1

3. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + u_i$.

a. Show that the OLS estimator $\hat{\beta}_1$ can be written as $\hat{\beta}_1 = r_{x,y}\frac{s_y}{s_x}$, where $r_{x,y}$ is the sample correlation between x and y, and s_y and s_x are the sample standard deviations of y and x, respectively.

$$r_{x,y} = \frac{Cov(X,Y)}{\sqrt{VarX\, VarY}}, \quad S_Y = \sqrt{VarY}, \quad S_X = \sqrt{VarX}$$

$$\Rightarrow r_{x,y}\frac{S_Y}{S_X} = \frac{Cov(X,Y)}{VarX} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \hat{\beta}_1$$

b. Show that the OLS estimator $\hat{\beta}_0$ is an unbiased estimator of β_0. (Hint: Use the fact that $\hat{\beta}_1$ is an unbiased estimator of β_1.)

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = (\beta_0 + \beta_1\bar{X} + \bar{u}) - \hat{\beta}_1\bar{X}$$

$$= \beta_0 + \bar{u} + (\beta_1 - \hat{\beta}_1)\bar{X}$$

$$E(\hat{\beta}_0|X) = E(\beta_0 + \bar{u} + (\beta_1-\hat{\beta}_1)\bar{X}|X)$$

$$= \beta_0 + E(\bar{u}|X) + \bar{X}\,\underline{E(\beta_1-\hat{\beta}_1|X)}$$

$$= \beta_0$$

$E(\hat{\beta}_1) = \beta_1 \Rightarrow E(\beta_1-\hat{\beta}_1|X)=0$

$\{(X_i, Y_i), i=1,\dots,n\}$ is random sample

$$\Rightarrow E(\hat{\beta}_0) = E(E(\hat{\beta}_0|X)) = \beta_0.$$

$\Rightarrow$ $\hat{\beta}_0$ is an unbiased esimator of β_0.

4. Use the data in ATTEND.TXT for this exercise.

(i) Obtain the minimum, maximum, and average values for the variables $atndrte$, $priGPA$, and ACT, where $atndrte$ is percent classes attended, $priGPA$ is cumulative GPA prior to term, and ACT is the ACT score.

	min	max	average
atndrte	6.25	100	81.70965
priGPA	0.857	3.93	2.586775
ACT	13	32	22.51029

(ii) Estimate the model

$$atndrte = \beta_0 + \beta_1\, priGPA + \beta_2\, ACT + u,$$

and write the results in equation form. Interpret the intercept. Does it have a useful meaning?

$atndrte = 75.700 + 17.261\, priGPA - 1.717\, ACT + u.$

intercept 75.700 is the expected value when priGPA = 0 and ACT = 0.

75.700 is the prediction of class attended percent when prior GPA and ACT are zero.

(iii) Interpret the estimated slope coefficients. Are there any surprises?

17.261 : the expected increase of class attended percent is 17.261 when prior GPA increases 1.

-1.717 : the expected decrease of class attended percent is 1.717 when ACT increases 1.

it is surprising that the slope coefficient of ACT is negative.

(iv) What is the predicted *atndrte* if *priGPA* = 3.65 and *ACT* = 20? What do you make of this result? Are there any students in the sample with these values of the explanatory variables?

104.3705

it exceeds the limit 100

which is impossible.

Yes

(v) If Student A has *priGPA* = 3.1 and *ACT* = 21 and Student B has *priGPA* = 2.1 and *ACT* = 26, what is the predicted difference in their attendance rates?

predicted atndrte of A is 93.16063

of B is 67.31727

difference is 25.84336.

5. Use the data in HPRICE1.TXT to estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

where *price* is the house price in thousands of dollars.

(i) Write out the results in equation form.

$price = -19.3150 + 0.1284\, sqrft + 15.1982\, bdrms + u.$

(ii) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

15.1982

(iii) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in part (ii).

$15.1982 + 0.1284 \times 140 = 33.1742$

(iv) What percentage of variation in price is explained by square footage and number of bedrooms?

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum u_i^2}{\sum (y_i - \bar{y})^2} = 0.6319184$$

$\Rightarrow$ 63.1918% is explained.

(v) The first house in the sample has *sqrft* = 2,438 and *bdrms* = 4. Find the predicted selling price for this home from the OLS regression line.

354.6052

(vi) The actual selling price of the first house in the sample was \$300,000 (so *price* = 300). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

−54.60525

No. it is the error of the model.

6. Use the data set in WAGE2.TXT for this problem. As usual, be sure all of the following regressions contain an intercept.

(i) Run a simple regression of *IQ* on *educ* to obtain the slope coefficient, say, $\tilde{\delta}_1$.

$$\overline{IQ} = 53.687 + 3.534\,\overline{educ}$$

$$\underline{\tilde{\delta}_1 = 3.534}$$

(ii) Run the simple regression of log(*wage*) on *educ*, and obtain the slope coefficient, $\tilde{\beta}_1$.

$$\overline{\log(wage)} = 5.97306 + 0.05984\,\overline{educ}$$

$$\tilde{\beta}_1 = 0.05984$$

(iii) Run the multiple regression of log(*wage*) on *educ* and *IQ*, and obtain the slope coefficients, $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively.

$$\overline{\log(wage)} = 5.658288 + 0.039120\,\overline{educ} + 0.005863\, IQ$$

$$\hat{\beta}_1 = 0.039120 \qquad \hat{\beta}_2 = 0.005863$$

(iv) Verify that $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$.

$$\hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1 = 0.05983984 \approx 0.05984 = \tilde{\beta}_1$$

$$\Longrightarrow \tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$$