EC3303 Econometrics I Tutorial Problem Set 2

1. In Singapore, only selected JC and Poly students take the SATs after their "A" levels or diploma. Many of them take the SATs because they would like to attend universities in North America. So the lot who take the SATs are likely to be a "positively selected" lot (i.e. they are likely to be smarter and more motivated than the average JC or Poly student).
   Suppose the mean SAT math score among those that took the SAT is 519 in Singapore. And further suppose that your friend claims that if *all* JC and Poly students in Singapore were made to take the test, then the mean score would be 450. You give the test to a random sample of 500 JC and Poly students in Singapore and found that these students had a mean score of 461. Is this good evidence against your friend's claim? Assume $s_Y = 100$

   Answer: $H_0: E(Y) = 450;\ \ H_1: E(Y) \neq 450$
   The t-statistic is $t^{act} = \frac{\bar{Y}^{act} - \mu_{Y,0}}{s_Y/\sqrt{n}} = \frac{461-450}{100/\sqrt{500}} = 2.46$
   The p-value is $2 \times Pr(t \geq 2.46) = 2 \times \Pr(t \leq -2.46) = 0.0138$
   So we reject the claim that the mean for all JC and Poly students in Singapore is equal to 450 at the 5% level (note conclusion will depend on the significance level chosen).

2. Suppose that a researcher, using data on class size(CS) and average test scores from 100 primary 3 classes, estimates the OLS regression

$$(\widehat{TestScore} = 731.4 - 3.42 \times CS, \qquad R^2 = 0.10, \qquad SER = 11.0)$$

a. A classroom has 19 students. What is the regression's prediction for that classroom's average test score?
b. Last year, a classroom had 17 students and this year it has 22 students. What is the regression's prediction for the change in the classroom average test score?
c. The sample average class size across the 100 classroom is 21.4. What is the sample average of the test scores across the 100 classrooms?

Answer:

a. The predicted average test score is

$$\widehat{Testscore} = 731.4 - 3.42 \times 19 = 666.4$$

b. The predicted change in the classroom test score is

$$\Delta\widehat{Testscore} = -3.42 \times 22 - (-3.42 \times 17) = -75.24 + 58.14 = -17.1$$

c. We know that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$. So The sample average of the test scores across the 100 classroom is

$$\overline{Testscore} = \hat{\beta}_0 + \hat{\beta}_1 \times \overline{CS} = 731.4 - 3.42 \times 21.4 = \ 658.2$$

3. You are interested in examining the relationship between earnings and height. Accordingly you run a regression of *Earn* on *Height* using a sample of American workers (where the variable *Earn* represents annual labour earnings in US dollars in 2015; and where *Height* represents the height of the worker in inches in 2015). The height of individuals in your sample ranges from 48 inches to 84 inches.

   You obtained the following regression output:

```
. regress  Earn Height, robust

Linear regression                               Number of obs =    17870
                                                F(  1, 17868) =   197.19
                                                Prob > F      =   0.0000
                                                R-squared     =   0.0109
                                                Root MSE      =    26777

------------------------------------------------------------------------------
             |               Robust
        Earn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Height |   707.6716    50.39502    14.04   0.000     608.8924    806.4507
       _cons |  -512.7336    3379.864    -0.15   0.879    -7137.594    6112.126
------------------------------------------------------------------------------
```

For all questions below, provide your **final answers** to 2 decimal places.

a. How much more or less do we expect a person whose height is 72 inches to earn in annual earnings compared to a person whose height is only 62 inches?

b. Suppose a person, Jane, actually earns \$33,712.97 per week. Jane is 59 inches tall. How large is the residual specific for Jane?

Answer:

a. A person whose height is 72 inches is predicted to earn $707.6716(72 - 62) = \$7{,}076.72$ more in annual earnings than a person whose height is 62 inches.

b. Jane's predicted annual earnings is $-512.7336 + 707.6716 \times 59 = \$41{,}239.89$. The residual specific for Jane is \$33,712.97 - \$41,239.89 = - \$7,526.92.

<u>Stata Exercise (to be done in tutorial with the tutor)</u>

4. The data file CPS08.dta contains data for full-time workers, aged 25-34, with a high school diploma (equivalent to Secondary, JC, and Poly qualifications) or Bachelor's as their highest degree. In this exercise, you will investigate the relationship between a worker's age and earnings (generally, older workers have more job experience, leading to higher productivity and earnings)

a. Run a regression of average hourly earnings (AHE) on age (Age). What is the estimated intercept? What is the estimated slope? Use the estimated regression to answer this question: How much do earnings increase as workers age by 1 year?

b. Ah Teck is a 26 year-old worker. Predict Ah Teck's earnings using the estimated regression. Ravin is a 30-year old worker. Predict Ravin's earnings using the estimated regression.

c. Does age account for a large fraction of the variance in earnings across individuals? Explain.

Answer:

a. $\widehat{AHE} = 1.08 + 0.60 \times Age$
Earnings increase, on average, by 0.6 dollars per hour when workers age by 1 year.

b. Ah Teck's predicted earnings = $1.08 + (0.60 \times 26) = \$16.68$
Ravin's predicted earnings = $1.08 + (0.60 \times 30) = \$19.08$

c. The regression R-squared is 0.03. This means that age explains a small fraction of the variability in earnings across individuals.