EC3303 Econometrics I Tutorial Problem Set 2

1. You have read in the news that women earn a lower salary than men on average. To test this hypothesis, you collect a random sample of 268 individuals and regressed weekly earnings (measured in $) on a constant and a binary variable, which takes on a value of 1 for females and is 0 otherwise. The results were (with standard errors in parentheses):

$$\widehat{Earn} = \underset{(9.44)}{578.70} - \underset{(13.52)}{169.72} \times Female,\ R^2 = 0.084,\ SER = 282.12.$$

a. What is the average weekly earnings for females in your sample? What is the difference between the average weekly earnings of men and women in your sample? Write your answer to 2 decimal places.

b. Are the population mean weekly earnings for men and women different at the 5% significance level? Explain how you obtained your conclusion.

c. Your colleague looks at the regression results you have obtained and concludes that there is discrimination against females. Do you agree or disagree with him? Explain.

Answer:

a. average weekly earnings for females in the sample = 578.70 - 169.72 = $408.98. Difference between the average weekly earnings of men and women in the sample = $169.72.

b. H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$

1) Compute the t-statistic: $t^{act} = \frac{\widehat{\beta^{act}}_1 - \beta_{1,0}}{SE(\widehat{\beta}_1)} = \frac{-169.72-0}{13.52} = -12.55$

2) $|t^{act}| = 12.55 > 1.96$, so we reject H_0 at the 5% significance level. The population mean earnings between men and women are different at the 5% level.

c. Differences in attributes of the individuals, such as education, ability, and tenure with an employer, have not been taken into account. Hence, in itself, this is weak evidence, at best, for discrimination.

2. An American researcher provides you with measurements of height in inches of 500 female and 500 male students (*Studenth*) at her university. A regression of *Studenth* on a binary variable (*BFemal*), which takes a value of one for females and is zero otherwise, yields the following result:

$$\widehat{Studenth} = \underset{(0.3)}{71.0} - \underset{(0.57)}{4.84} \times BFemal,\ R^2 = 0.40$$

a. What is the interpretation of the "intercept"? What is the interpretation of the "slope"?

b. Is there evidence at the 1% significance level that the mean heights of males and the mean heights of females are statistically significantly different?

c. How would your OLS regression line change if you were asked to report student heights in centimeters? (1 inch = 2.54cm)

Answer:

a. The intercept gives you the average height of males, which is 71.0 inches in this sample. The slope tells you by how much shorter females are, on average (4.84 inches), compared to males. The average height of females in the sample is therefore about 66.2 inches.

b. The *t*-statistic for the difference in means is -4.84/0.57= -8.49. Since $|t - statistic| = 8.49 > \ 2.58$, the mean heights of males and females are statistically significantly different at the 1% level.

c. The OLS regression line would become:

$$\widehat{Studenth} = 180.34 - 12.29 \times BFemal$$
$$(0.762) \quad (1.448)$$

3. Suppose you ran a simple regression of the average hourly earnings (*AHE*) of individuals on a binary variable indicating whether or not the individual is female (*Female*). The results were (with standard errors in parentheses):

$$\widehat{AHE} = 20.11 - 2.63 \times Female,\ R^2=0.0165,\ SER = 10.06.$$
$$(0.16) \quad (0.23)$$

Further, suppose that in your sample, the standard deviation of hourly earnings for men were 10.68 while the standard deviation of hourly earnings for women were 9.17. Would you expect the homoskedasticity assumption to hold in this setting? Explain carefully why or why not.

Answer: In this application, the variance of hourly earnings for men is 10.68^2= 114.1 while variance of hourly earnings for women is 9.17^2= 84.1.

If the population errors are homoskedastic, $Var(u_i|X_i = x) =$constant. That is, the variance of the conditional distribution of the error terms given any value of *X* does not vary with *X*.

In this application, u_i is the deviation of the i^{th} man's earnings from the population mean earnings for men (β_0) and the deviation of the i^{th} woman's earnings from the population mean earnings for women ($\beta_0 + \ \beta_1$). So the statement "the variance of u_i does not vary with *X* is equivalent to the statement "the variance of weekly earnings is the same for men as it is for women".

Since the spread of earnings for men is likely to be greater than the spread of earnings for women, the population errors are unlikely to be homoskedastic in this case.

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

4. The California Standardized Testing and Reporting data set (caschool.dta) contains data on test performance, school characteristics, and student demographic backgrounds for 420 Californian school districts. Today, we will learn how to perform a simple linear regression and linear regression with binary regressors.