Moh JAISWAL
moh.jaiswal@ryerson.ca
500916860

Assignment 3

5.2 Suppose that a researcher, using wage data on 250 randomly selected male workers and 280 female workers, estimates the OLS regression

$$\widehat{Wage} = 12.52 + 2.12 \times Male, R^2 = 0.06, SER = 4.2,$$
$$(0.23) \quad (0.36)$$

where *Wage* is measured in dollars per hour and *Male* is a binary variable that is equal to 1 if the person is a male and 0 if the person is a female. Define the wage gender gap as the difference in mean earnings between men and women.

a. What is the estimated gender gap?

b. Is the estimated gender gap significantly different from 0? (Compute the *p*-value for testing the null hypothesis that there is no gender gap.)

c. Construct a 95% confidence interval for the gender gap.

d. In the sample, what is the mean wage of women? Of men?

e. Another researcher uses these same data but regresses *Wages* on *Female*, a variable that is equal to 1 if the person is female and 0 if the person a male. What are the regression estimates calculated from this regression?

$$\widehat{Wage} = \_\_\_ + \_\_\_ \times Female, R^2 = \_\_\_, SER = \_\_\_.$$

a) Earnings for men : $12.52 + 2.12 \times 1 = 14.64$

" " women : $12.52 + 2.12 \times 0 = 12.52$

Estimated gender gap : $14.64 - 12.52 = 2.12$

b) $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ such that $\beta_1 = \mu_m - \mu_f$

$$t^{stat} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{2.12 - 0}{0.36} = 5.89$$

$$p\text{-val} = 2\Phi(-|t^{stat}|) = 2 \times 0 \times -5.89 = 0$$

p-val < 0.05 so we reject null hypothesis

c) At 95% CVal $= \pm 1.96$

C.F. $\Rightarrow (2.12 \pm 1.96 \times 0.36)$

$1.41 \leq \beta_1 \leq 2.83$

d) Mean wage for men $= 14.64$/hr
" " " women $= 12.52$/hr

e) Old eqn:

Mean wage for men $= \beta_0 + \beta_1$
" " " women $= \beta_0$

New eqn $\Rightarrow$ Wage $= a_0 + a_1 \times$ female

a_0 is avg. wage for men

$a_0 = \beta_0 + \beta_1 = 14.64$

$a_0 + a_1$ = avg. wage for women $= \beta_0 \therefore a_1 = \beta_0 - a_0$

$\therefore$ new eqn is:

$\beta_0 + \beta_1 + \beta_0 \times$ Female

$\Rightarrow$ $14.64 - 2.12 \times$ female is the new eqn.

Since residuals remain unchanged, so do R^2 & SER val.

Wage $= 14.64 - 2.12 \times$ Female; $R^2 = 0.06$, SER $= 4.2$

5.5 In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to "regular" and "small" classes and given standardized tests at the end of the year. (Regular classes contained approximately 24 students, and small classes contained approximately 15 students.) Suppose that, in the population, the standardized tests have a mean score of 925 points and a standard deviation of 75 points. Let *SmallClass* denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise. A regression of *TestScore* on *SmallClass* yields

$$\widehat{TestScore} = 918.0 + 13.9 \times SmallClass,\ R^2 = 0.01,\ SER = 74.6.$$
$$(1.6) \quad (2.5)$$

a. Do small classes improve test scores? By how much? Is the effect large? Explain.

b. Is the estimated effect of class size on test scores statistically significant? Carry out a test at the 5% level.

c. Construct a 99% confidence interval for the effect of *SmallClass* on *Test Score*.

a) Improves scores by 13.9 points. It is one fifth (18.6%) of SER and therefore a conservative increase and not very large.

b) $H_0 : \beta_1 = 0 \ ; \ H_1 : \beta_1 \neq 0$

$$t^{stat} = \frac{13.9 - 0}{2.5} = 5.56$$

$$p\text{-val} = 2\Phi(-|t^{stat}|) = Pr(|z| > |t^{stat}|)$$

$p\text{-val} = 0$

∴ reject null hypothesis.

c)

cv at 99% Confidence $= \pm 2.58$

C.I. $\Rightarrow (13.9 \pm 2.58 \times 2.5)$

$$7.45 \leq \beta_1 \leq 20.35$$

5.8 Suppose that (Y_i, X_i) satisfy the least squares assumptions in Key Concept 4.3 and, in addition, u_i is $N(0, \sigma_u^2)$ and is independent of X_i. A sample of size $n = 30$ yields

$$\underset{(10.2)}{\hat{Y} = 43.2 + 61.5X}, R^2 = 0.54, SER = \underset{(7.4)}{1.52},$$

where the numbers in parentheses are the homoskedastic-only standard errors for the regression coefficients.

a. Construct a 95% confidence interval for β_0.

b. Test H_0: $\beta_1 = 55$ vs. H_1: $\beta_1 \neq 55$ at the 5% level.

c. Test H_0: $\beta_1 = 55$ vs. H_1: $\beta_1 > 55$ at the 5% level.

a)

Cval @ 95% confidence = ± 2.05

$\therefore$ $22.29 \leq \beta_1 \leq 64.11$

b) $t^{stat} \Rightarrow \dfrac{61.5 - 55}{7.4} = 0.88$

since $2.05 > 0.88$ we do not reject null hypothesis

c) Cval at 5% one tailed is 1.7

since $1.70 > 0.88$ we do not reject null hypothesis.

Exercises

The first four exercises refer to the table of estimated regressions on page 209, computed using data for 2012 from the CPS. The data set consists of information on 7440 full-time, full-year workers. The highest educational achievement for each worker was either a high school diploma or a bachelor's degree. The workers' ages ranged from 25 to 34 years. The data set also contains information on the region of the country where the person lived, marital status, and number of children. For the purposes of these exercises, let

AHE = average hourly earnings (in 2012 dollars)
College = binary variable (1 if college, 0 if high school)
Female = binary variable (1 if female, 0 if male)
Age = age (in years)
Ntheast = binary variable (1 if Region = Northeast, 0 otherwise)
Midwest = binary variable (1 if Region = Midwest, 0 otherwise)
South = binary variable (1 if Region = South, 0 otherwise)
West = binary variable (1 if Region = West, 0 otherwise)

6.1 Compute $\overline{R}^2$ for each of the regressions.

Results of Regressions of Average Hourly Earnings on Gender and Education Binary Variables and Other Characteristics, Using 2012 Data from the Current Population Survey

Dependent variable: average hourly earnings (AHE).			
Regressor	**(1)**	**(2)**	**(3)**
College (X_1)	8.31	8.32	8.34
Female (X_2)	−3.85	−3.81	−3.80
Age (X_3)		0.51	0.52
Northeast (X_4)			0.18
Midwest (X_5)			−1.23
South (X_6)			−0.43
Intercept	17.02	1.87	2.05
Summary Statistics			
SER	9.79	9.68	9.67
R^2	0.162	0.180	0.182
$\overline{R}^2$			
n	7440	7440	7440

1) $\overline{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$

$= 1 - \left(\frac{7440-1}{7440-2-1}\right) \times (1-0.162) = 0.16177 = 0.162$

(1)

(2) $= 0.180$

(3) $= 0.181 = 1 - \left(\frac{7440-1}{7440-6-1}\right) \times (1-0.182)$

6.2 Using the regression results in column (1):

a. Do workers with college degrees earn more, on average, than workers with only high school degrees? How much more?

b. Do men earn more than women, on average? How much more?

6.3 Using the regression results in column (2):

a. Is age an important determinant of earnings? Explain.

b. Sally is a 29-year-old female college graduate. Betsy is a 34-year-old female college graduate. Predict Sally's and Betsy's earnings.

6.4 Using the regression results in column (3):

a. Do there appear to be important regional differences?

b. Why is the regressor *West* omitted from the regression? What would happen if it were included?

2)a) They earn $8.31/hr more.

b) Women earn 3.85/hr less than men.

3)a) With an increase by a year they earn $0.51/hr more.

b) Sally's earnings = 1.87 + 8.32 − 3.81 + 0.51 × 29 = $21.17/hr
Betsy's earns 0.51 × 5 $/hr more = $23.72/hr

4)a) Northeast earns $0.18/hr more than the west.
Midwest earn $1.23/hr less than the west.
South earns $0.43/hr less than the west.

b) It is to avoid perfect multicollinearity.
If it were included, then the intercept would be written as a linear function of the four regional regressors.

c. Juanita is a 28-year-old female college graduate from the South. Jennifer is a 28-year-old female college graduate from the Midwest. Calculate the expected difference in earnings between Juanita and Jennifer.

c) $-0.43 - (-1.23) = +.80$/hr difference

Juanita earns 80¢ more/hr.

6.6 A researcher plans to study the causal effect of police on crime, using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police force.

a. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?

b. Use your answer to (a) and the expression for omitted variable bias given in Equation (6.1) to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. (That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?)

a) For omitted variable bias to occur the included regressor must be correlated to the omitted variable. There are other important determinants of a county's crime rate; we could add demographic factors such as age and gender to control for important omitted variables. Eg: Young males

b) Using the age and sex demographics from (a), we assume crime rate is positively affected by the fraction of young males to total population. The size of police force is directly correlated with the size of young males leading to the regression over-estimating the effect of police force on crime rate. Thus $\hat{\beta}_1 > \beta_1$

6.9 (Y_i, X_{1i}, X_{2i}) satisfy the assumptions in Key Concept 6.4. You are interested in β_1, the causal effect of X_1 on Y. Suppose that X_1 and X_2 are uncorrelated. You estimate β_1 by regressing Y onto X_1 (so that X_2 is not included in the regression). Does this estimator suffer from omitted variable bias? Explain.

Since X_1 and X_2 are uncorrelated $\hat{\beta}_1$ does not suffer from omitted variable bias.