ECO 321.03: Homework 3

Due on 10/09

Maria Perez-Urdiales

Harris Temuri

Problem 1

Let us consider the following regression model for a sample of n female workers

$$\textbf{WorkHours}_i = \beta_0 + \beta_1 \textbf{Child}_i + u_i,\ i = 1, ..., n, \tag{1}$$

where $Child_i$ is a binary variable that takes value 1 if the individual has one or more child and 0 otherwise; and $WorkHours$ is the individual's usual hours worked per week in past 12 months.
Let **WeeklyPay** $= Y$ and **Child** $= X$. Also let n_1 the number of individuals who has a one or more than one child and n_0 the number of female worker who does not have a child, such that $n_0 + n_1 = n$.
Recall that the OLS estimators are given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X})}{\sum_{i=1}^{n}(x_i - \bar{X})^2}$$
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

a. Let $y_{i,0}$ an individual i's working hours for which the value of X is 0; and $y_{i,1}$ an individual i's working hour for which the value of X is 1. Show that
$$\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X}) = \frac{n_0}{n}\sum_{i=1}^{n_0} y_{i,1} - \frac{n_1}{n}\sum_{i=1}^{n_0} y_{i,0}$$

b. Show that
$$\sum_{i=1}^{n}(x_i - \bar{X})^2 = \frac{n_1 n_0}{n}$$

c. Use the results in (a) and (b) to conclude that
$$\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$$
upon an appropriate definition of $\bar{Y}_1$ and $\bar{Y}_0$.

d. Show that $\bar{Y} = (n_1/n)\bar{Y}_1 + (n_0/n)\bar{Y}_0$

e. Using the result in (d), finally show that $\hat{\beta}_0 = \bar{Y}_0$

Solution

Part A

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$
$$y_{i,0} = \beta_0 + u_i$$
$$y_{i,1} = \beta_0 + \beta_1 + u_i$$
$$\beta_1 = E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$$
$$\beta_1 = \frac{\sum_{i=1}^{n_1} y_{i,1}}{n_1} - \frac{\sum_{i=1}^{n_0} y_{i,0}}{n_0}$$
$$\frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X})}{\sum_{i=1}^{n}(x_i - \bar{X})^2} = \frac{\sum_{i=1}^{n_1} y_{i,1}}{n_1} - \frac{\sum_{i=1}^{n_0} y_{i,0}}{n_0}$$
$$\frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X})}{\frac{n_1 n_0}{n}} = \frac{\sum_{i=1}^{n_1} y_{i,1}}{n_1} - \frac{\sum_{i=1}^{n_0} y_{i,0}}{n_0}$$
$$\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X}) = \frac{n_0}{n}\sum_{i=1}^{n_0} y_{i,1} - \frac{n_1}{n}\sum_{i=1}^{n_0} y_{i,0}$$

Problem 1 continued on next page. . .

Part B

$$\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X}) = \frac{n_0}{n}\sum_{i=1}^{n_0} y_{i,1} - \frac{n_1}{n}\sum_{i=1}^{n_0} y_{i,0}$$

$$\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X}) = \frac{n_1 n_0}{n}\left(\frac{\sum_{i=1}^{n_1} y_{i,1}}{n_1} - \frac{\sum_{i=1}^{n_0} y_{i,0}}{n_0}\right)$$

$$\frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X})}{\frac{n_1 n_0}{n}} = \frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X})}{\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

$$\frac{n_1 n_0}{n} = \sum_{i=1}^{n}(x_i - \bar{X})^2$$

Part C

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n_1} y_{i,1}}{n_1} - \frac{\sum_{i=1}^{n_0} y_{i,0}}{n_0}$$

$$\hat{\beta}_1 = \frac{n_1 \bar{Y}_1}{n_1} - \frac{n_0 \bar{Y}_0}{n_0}$$

$$\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$$

Part D

$$\bar{Y} = \frac{\sum_{i=0}^{n} y_i}{n}$$

$$\bar{Y} = \frac{\sum_{i=1}^{n_1} y_{i,1} + \sum_{i=1}^{n_0} y_{i,0}}{n}$$

$$\bar{Y} = \frac{n_1 \bar{Y}_1 + n_0 \bar{Y}_0}{n}$$

$$\bar{Y} = (n_1/n)\bar{Y}_1 + (n_0/n)\bar{Y}_0$$

Part E

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_0 = \bar{Y}_0 - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_0 = \bar{Y}_0 - \hat{\beta}_1(0)$$

$$\hat{\beta}_0 = \bar{Y}_0$$

Problem 2

The data file, **female work.csv**, is a subset of The American Community Survey(ACS) for the year 2018, with only female workers from age 18-40. A detailed description of the variables contained in the file is given in the pdf file **Documentation.pdf** available on Blackboard. You can find a sample code for fitting and testing a linear regression model in R inside the folders called "Chapter 2: Linear Regression with One Regressor" and "Chapter 3: Inference in the Linear Model with One Regressor" on Blackboard.

a. Suppose that all assumptions for OLS are satisfied and estimate the simple regression model in equation (1) using heteroskedasticity robust standard errors.

b. Report the values for $\hat{\beta}_0$ and $\hat{\beta}_1$.

c. What does the sample statistic $\hat{\beta}_0$ capture?

d. Do women with child work less? By how much? Explain.

e. Is the estimated effect of having child on women's working hours statistically significant? Carry out a test at the 1% level.

f. Construct a 95% confidence interval for the effect of having children on working hours.

Solution

Part A

```
model <- lm(workhour ~ child, data = FemaleWork)
coeftest(model, vcov = vcovHC(model, type = "HC1"))
# t test of coefficients:
#
#              Estimate Std. Error  t value  Pr(>|t|)
# (Intercept) 36.250973   0.028318 1280.143 < 2.2e-16 ***
# child       (neg)1.009660   0.041918  -24.087 < 2.2e-16 ***
# ---
# Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

$$\textbf{WorkHours}_i = \beta_0 + \beta_1 \textbf{Child}_i + u_i,$$
$$\textbf{WorkHours}_i = 36.25 - 1.01 \textbf{Child}_i + u_i,$$

Part B

The value for $\hat{\beta}_0$ is 36.25 with $SE(\hat{\beta}_0) = 0.028$. The value for $\hat{\beta}_1$ is -1.01 with $SE(\hat{\beta}_1) = 0.042$.

Part C

It's an estimate on how many hours a female without a child works.

Part D

Women with one or more children tend to work 1.01 fewer hours than women without children.

Problem 2 continued on next page. . .

Part E

```r
summary(model)

# Call:
# lm(formula = workhour ~ child, data = FemaleWork)
#
# Residuals:
#     Min      1Q  Median      3Q     Max
# -35.251  -6.251   3.749   4.759  63.759
#
# Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) 36.25097    0.02829 1281.50   <2e-16 ***
# child       -1.00966    0.04193  -24.08   <2e-16 ***
# ---
# Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
#
# Residual standard error: 11.96 on 328397 degrees of freedom
# Multiple R-squared:  0.001763,  Adjusted R-squared:  0.00176
# F-statistic: 579.9 on 1 and 328397 DF,  p-value: < 2.2e-16
```

The p-value for the linear model is much less than 0.01 so the model is statistically significant.

Part F

```r
newdata <- data.frame(child=1)
predict(model, newdata = newdata, interval = "confidence")
#        fit      lwr      upr
#  35.24131 35.18066 35.30196
```

The 95% confidence interval for the effect of having children on working hours is $[35.18066, 35.30196]$.