TA Session 5

Chi-Yuan Fang

March 30, 2021

Contents

1 Introduction

1.1 TA Information

TA: Chi-Yuan Fang

TA sessions: Tuesday 1:20 – 3:10 PM (SS 501)

Email: r09323017@ntu.edu.tw

Office hours: Friday 2:00 – 3:30 PM or by appointments (SS 643)

Class group on Facebook: Statistics (Fall 2020) and Econometrics (Spring 2021)

https://www.facebook.com/groups/452292659024369/

Because screens are not clear in SS 501, I will provide the link of live streaming in the group.

1.2 TA Sessions Schedule

Week	TA Sessions	Quiz	Content	Remind
1	02/23: No class			
2	03/02: Class 1		Function, Confidence Interval, T test	03/10 Turn in HW1
3	03/09: Class 2		Loops, Linear Model	03/10 Turn in HW1, 03/16 Quiz 1
4	03/16: Class 3	Quiz 1	OLS	03/24 Turn in HW2
5	03/23: Class 4		Multiple Regression	03/24 Turn in HW2, 03/30 Quiz 2
6	03/30: Class 5	Quiz 2	Omitted Variable, F test	04/14 Turn in HW3
7	04/06: No class			04/14 Turn in HW3
8	04/13: Class 6			04/14 Turn in HW3, 04/20 Quiz 3
9	04/20: Class 7	Quiz 3		**04/28 Midterm**
10	04/27: Class 8		Review and Q&A	**04/28 Midterm**, 05/05 Turn in HW4
11	05/04: Class 9			05/05 Turn in HW4, 05/11 Quiz 4
12	05/11: Class 10	Quiz 4		05/19 Turn in HW5

Week	TA Sessions	Quiz	Content	Remind
13	05/18: Class 11			05/19 Turn in HW5, 05/25 Quiz 5
14	05/25: Class 12	Quiz 5		06/02 Turn in HW6
15	06/01: Class 13			06/02 Turn in HW6, 06/08 Quiz 6
16	06/08: Class 14	Quiz 6	Review and Q&A	**06/16 Final Exam**
17	06/15: No class			**06/16 Final Exam**
18	06/22: No class			

1.3 Reference

Introduction to Econometrics with R

https://www.econometrics-with-r.org

R for Data Science

https://r4ds.had.co.nz

R Markdown

https://rmarkdown.rstudio.com

Introduction to R Markdown

https://rpubs.com/brandonkopp/RMarkdown

What is a good book on learning R with examples?

https://www.quora.com/What-is-a-good-book-on-learning-R-with-examples

2 Empirical Exercise 7.1

Use the **Birthweight_Smoking** data set introduced in Empirical Exercise E5.3 to answer the following questions. To begin, run three regressions:

(1) *Birthweight* on *Smoker*
(2) *Birthweight* on *Smoker*, *Alcohol*, and *Nprevist*
(3) *Birthweight* on *Smoker*, *Alcohol*, *Nprevist*, and *Unmarried*

a. What is the value of the estimated effect of smoking on birth weight in each of the regressions?

Solution

```
library(kableExtra); library(stargazer)
library(ggplot2); library(dplyr);
library(jtools); library(ggstance);
library(broom.mixed); library(huxtable)
```

```
# import data set
library(readxl)
Birthweight_Smoking <- read_excel("Birthweight_Smoking/Birthweight_Smoking.xlsx")

# variable names
colnames(Birthweight_Smoking)
```

```
##  [1] "nprevist"    "alcohol"     "tripre1"     "tripre2"     "tripre3"
##  [6] "tripre0"     "birthweight" "smoker"      "unmarried"   "educ"
## [11] "age"         "drinks"
```

```r
# https://cran.r-project.org/web/packages/stargazer/stargazer.pdf
# https://www.jakeruss.com/cheatsheets/stargazer/
# https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf

library(sandwich); library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
# Model 1
E71_M1 <- lm(formula = birthweight ~ smoker,
             data = Birthweight_Smoking)

# Model 2
E71_M2 <- lm(formula = birthweight ~ smoker + alcohol + nprevist,
             data = Birthweight_Smoking)

# Model 3
E71_M3 <- lm(formula = birthweight ~ smoker + alcohol + nprevist + unmarried,
             data = Birthweight_Smoking)

# Adjust standard errors
cov1 <- vcovHC(E71_M1, type = "HC1")
rb_se1 <- sqrt(diag(cov1))

cov2 <- vcovHC(E71_M2, type = "HC1")
rb_se2 <- sqrt(diag(cov2))

cov3 <- vcovHC(E71_M3, type = "HC1")
rb_se3 <- sqrt(diag(cov3))

# output table
stargazer(E71_M1, E71_M2, E71_M3,
          type = "text",
          se = list(rb_se1, rb_se2, rb_se3),
          digits = 4)
```

```
##
## ==================================================================================================
##                                                    Dependent variable:
##                     ------------------------------------------------------------------------------
##                                                       birthweight
##                                  (1)                       (2)                       (3)
## --------------------------------------------------------------------------------------------------
## smoker                      -253.2284***              -217.5801***              -175.3769***
##                               (26.8104)                 (26.1076)                 (26.8268)
##
## alcohol                                                  -30.4913                  -21.0835
##                                                         (72.5967)                 (72.9921)
```

```
##
## nprevist                                              34.0699***               29.6025***
##                                                       (3.6083)                 (3.5827)
##
## unmarried                                                                    -187.1332***
##                                                                               (27.6772)
##
## Constant                  3,432.0600***          3,051.2490***           3,134.4000***
##                             (11.8905)              (43.7145)               (44.1486)
##
## ----------------------------------------------------------------------------------------------
## Observations                  3,000                  3,000                   3,000
## R2                            0.0286                 0.0729                  0.0886
## Adjusted R2                   0.0283                 0.0719                  0.0874
## Residual Std. Error   583.7297 (df = 2998)      570.4708 (df = 2996)      565.6975 (df = 2995)
## F Statistic         88.2793*** (df = 1; 2998) 78.4697*** (df = 3; 2996) 72.7930*** (df = 4; 2995)
## ==============================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```

b. Construct a 95% confidence interval for the effect of smoking on birth weight, using each of the regressions.

Solution

```r
# https://stats.stackexchange.com/questions/117052/replicating-statas-robust-option-in-r
library(estimatr)

# Model 1 # use HC1
E71_M1n <- lm_robust(formula = birthweight ~ smoker,
                     data = Birthweight_Smoking,
                     se_type = "stata")

# Model 2
E71_M2n <- lm_robust(formula = birthweight ~ smoker + alcohol + nprevist,
                     data = Birthweight_Smoking,
                     se_type = "stata")

# Model 3
E71_M3n <- lm_robust(formula = birthweight ~ smoker + alcohol + nprevist + unmarried,
                     data = Birthweight_Smoking,
                     se_type = "stata")

export_summs(E71_M1n, E71_M2n, E71_M3n,
             digits = 4,
             ci_level = 0.95,
             error_format = "[{conf.low}, {conf.high}]",
             model.names = c("Model (1)", "Model (2)", "Model (3)"))
```

c. Does the coefficient on Smoker in regression (1) suffer from omitted variable bias? Explain.

Solution

```r
(E71_M2$coefficients[2] - E71_M1$coefficients[2]) / E71_M1$coefficients[2]
```

```
##     smoker
## -0.1407752
```

	Model (1)	Model (2)	Model (3)
(Intercept)	3432.0600 ***	3051.2486 ***	3134.4000 ***
	[3408.7455, 3455.3744]	[2965.5352, 3136.9619]	[3047.8354, 3220.9646]
smoker	-253.2284 ***	-217.5801 ***	-175.3769 ***
	[-305.7970, -200.6597]	[-268.7708, -166.3894]	[-227.9777, -122.7761]
alcohol		-30.4913	-21.0835
		[-172.8357, 111.8531]	[-164.2032, 122.0363]
nprevist		34.0699 ***	29.6025 ***
		[26.9949, 41.1450]	[22.5777, 36.6274]
unmarried			-187.1332 ***
			[-241.4014, -132.8651]
nobs	3000	3000	3000
r.squared	0.0286	0.0729	0.0886
adj.r.squared	0.0283	0.0719	0.0874
statistic	89.2110	59.4841	56.0850
p.value	0.0000	0.0000	0.0000
df.residual	2998.0000	2996.0000	2995.0000
nobs.1	3000.0000	3000.0000	3000.0000
se_type	HC1.0000	HC1.0000	HC1.0000

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Yes, it does. The coefficient changes by roughly 14% in magnitude when additional regressors are added to (1). This change is substantively large and large relative to the standard error in (1).

d. Does the coefficient on Smoker in regression (2) suffer from omitted variable bias? Explain.

Solution

```
(E71_M3$coefficients[2] - E71_M2$coefficients[2]) / E71_M2$coefficients[2]
```

```
##     smoker
## -0.1939661
```

Yes, it does. The coefficient changes by roughly 19% in magnitude when *unmarried* is added as an additional regression. This change is substantively large and large relative to the standard error in (2).

e. Consider the coefficient on *Unmarried* in regression (3).
   i. Construct a 95% confidence interval for the coefficient.
   ii. Is the coefficient statistically significant? Explain.
   iii. Is the magnitude of the coefficient large? Explain.

iv. A family advocacy group notes that the large coefficient suggests that public policies that encourage marriage will lead, on average, to healthier babies. Do you agree? (Hint: Review the discussion of control variables in Section 6.8. Discuss some of the various factors that *Unmarried* may be controlling for and how this affects the interpretation of its coefficient.)

Solution

i. See part (b).

ii.

- **Prepare**

  $H_0 : \beta_{unmarried} = 0$ v.s. $H_1 : \beta_{unmarried} \neq 0$

  Let the significance level be 5%.

- **Calculate**

  See the results in part (a) or (b).

- **Conclusion**

  Because $p - value < \alpha = 5\%$, we reject H_0. There is significant evidence that the coefficient on *unmarried* is different from 0.

iii. Yes, it is. Other things being equal, *birthweight* is 187 grams lower for unmarried mothers on average.

iv. As the question suggests, *unmarried* is a control variable that captures the effects of several factors that differ between married and unmarried mothers such as age, education, income, diet and other health factors, and so forth.

f. Consider the various other control variables in the data set. Which do you think should be included in the regression? Using a table like Table 7.1, examine the robustness of the confidence interval you constructed in (b). What is a reasonable 95% confidence interval for the effect of smoking on birth weight?

Solution

We consider adding on additional regression in the table that includes *Age* and *Educ* (years of education) in model (4).

```r
# Model 4
E71_M4 <- lm(formula = birthweight ~ smoker + alcohol + nprevist + unmarried + age + educ,
             data = Birthweight_Smoking)

# Adjust standard errors
cov4 <- vcovHC(E71_M4, type = "HC1")
rb_se4 <- sqrt(diag(cov4))

# output table
stargazer(E71_M3, E71_M4,
          type = "text",
          column.labels=c("Model (3)","Model (4)"),
          se = list(rb_se3, rb_se4),
          digits = 4)
```

```
##
## ========================================================================
##                                          Dependent variable:
##                      ---------------------------------------------------
```

```
##                                       birthweight
##                            Model (3)                 Model (4)
##                               (1)                       (2)
## ----------------------------------------------------------------------
## smoker                    -175.3769***             -176.9589***
##                            (26.8268)                 (27.3314)
##
## alcohol                    -21.0835                  -14.7583
##                            (72.9921)                 (72.9074)
##
## nprevist                   29.6025***               29.7751***
##                             (3.5827)                 (3.5973)
##
## unmarried                 -187.1332***             -199.3195***
##                            (27.6772)                 (30.9943)
##
## age                                                   -2.4935
##                                                      (2.4451)
##
## educ                                                   0.2380
##                                                      (5.5328)
##
## Constant                 3,134.4000***            3,199.4260***
##                            (44.1486)                 (90.6361)
##
## ----------------------------------------------------------------------
## Observations                 3,000                     3,000
## R2                          0.0886                    0.0890
## Adjusted R2                 0.0874                    0.0872
## Residual Std. Error   565.6975 (df = 2995)      565.7600 (df = 2993)
## F Statistic         72.7930*** (df = 4; 2995) 48.7410*** (df = 6; 2993)
## ======================================================================
## Note:                                      *p<0.1; **p<0.05; ***p<0.01
```

The coefficient on *Smoker* in model (4) is very similar to its value in model (3).

3 Empirical Exercise 7.2

In the empirical exercises on earning and height in Chapters 4 and 5, you estimated a relatively large and statistically significant effect of a worker's height on his or her earnings. One explanation for this result is omitted variable bias: Height is correlated with an omitted factor that affects earnings. For example, Case and Paxson (2008) suggest that cognitive ability (or intelligence) is the omitted factor. The mechanism they describe is straightforward: Poor nutrition and other harmful environmental factors in utero and in early childhood have, on average, deleterious effects on both cognitive and physical development. Cognitive ability affects earnings later in life and thus is an omitted variable in the regression.

a. Suppose that the mechanism described above is correct. Explain how this leads to omitted variable bias in the OLS regression of *Earnings* on *Height*. Does the bias lead the estimated slope to be too large or too small? [Hint: Review Equation (6.1).]

Remark

Given true model

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + u. \tag{1}$$

Z is omitted variable if (1) $Cov(X,Z) \neq 0$, and (2) $\beta_2 \neq 0$. Fit the following models:

$$\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X \tag{2}$$
$$\widehat{Y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 X + \widetilde{\beta}_2 Z \tag{3}$$
$$\widehat{Z} = \widehat{\delta}_0 + \widehat{\delta}_1 X \tag{4}$$

where $\widetilde{\beta}_i$ is unbiased estimator of β_i, and $\widehat{\beta}_i$ is biased estimator of β_i. Thus,

$$\widehat{\beta}_1 = \widetilde{\beta}_1 + \widehat{\delta}_1 \times \widetilde{\beta}_2. \tag{5}$$

Solution

Because we know $Cov(height, cognitive\_ability) \neq 0$ and $Cov(cognitive\_ability, earning) \neq 0$, the simple regression model suffers from an omitted variables problem. Moreover, we think $Cov(height, cognitive\_ability) > 0$ and $Cov(cognitive\_ability, earning) > 0$, so $\widehat{\beta}_{height}$ is more positive than β_{height} (underestimated).

> If the mechanism described above is correct, the estimated effect of height on earnings should disappear if a variable measuring cognitive ability is included in the regression. Unfortunately, there isn't a direct measure of cognitive ability in the data set, but the data set does include years of education for each individual. Because students with higher cognitive ability are more likely to attend school longer, years of education might serve as a control variable for cognitive ability; in this case, including education in the regression will eliminate, or at least attenuate, the omitted variable bias problem.
>
> Use the years of education variable (educ) to construct four indicator variables for whether a worker has less than a high school diploma ($LT\_HS = 1$ if $educ < 12$, 0 otherwise), a high school diploma ($HS = 1$ if $educ = 12$, 0 otherwise), some college ($Some\_Col = 1$ if $12 < educ < 16$, 0 otherwise), or a bachelor's degree or higher ($College = 1$ if $educ \geq 16$, 0 otherwise).
>
> b. Focusing first on women only, run a regression of (1) $Earnings$ on $Height$ and (2) $Earnings$ on $Height$, including LT_HS, HS, and $Some\_Col$ as control variables.
>   i. Compare the estimated coefficient on $Height$ in regressions (1) and (2). Is there a large change in the coefficient? Has it changed in a way consistent with the cognitive ability explanation? Explain.
>   ii. The regression omits the control variable $College$. Why?
>   iii. Test the joint null hypothesis that the coefficients on the education variables are equal to 0.
>   iv. Discuss the values of the estimated coefficients on LT_HS, HS, and $Some\_Col$. (Each of the estimated coefficients is negative, and the coefficient on LT_HS is more negative than the coefficient on HS, which in turn is more negative than the coefficient on $Some\_Col$. Why? What do the coefficients measure?)

Solution

```
# https://cran.r-project.org/web/packages/stargazer/stargazer.pdf
# https://www.jakeruss.com/cheatsheets/stargazer/
# https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf

library(sandwich); library(lmtest)
library(stargazer)
library(car)
```

```
# import data
library(readxl)
Earnings_and_Height <- read_xlsx("Earnings_and_Height/Earnings_and_Height.xlsx")
```

```r
# create dummay variables
LT_HS <- c()

for (i in 1:length(Earnings_and_Height$educ)){
  if (Earnings_and_Height$educ[i] < 12){
    LT_HS[i] <- c(1)
  } else {
    LT_HS[i] <- c(0)
  }
}

HS <- c()

for (i in 1:length(Earnings_and_Height$educ)){
  if (Earnings_and_Height$educ[i] == 12){
    HS[i] <- c(1)
  } else {
    HS[i] <- c(0)
  }
}

Some_Col <- c()

for (i in 1:length(Earnings_and_Height$educ)){
  if (Earnings_and_Height$educ[i] > 12 & Earnings_and_Height$educ[i] < 16){
    Some_Col[i] <- c(1)
  } else {
    Some_Col[i] <- c(0)
  }
}


College <- c()

for (i in 1:length(Earnings_and_Height$educ)){
  if (Earnings_and_Height$educ[i] >= 16){
    College[i] <- c(1)
  } else {
    College[i] <- c(0)
  }
}

Earnings_and_Height <- cbind(Earnings_and_Height, LT_HS, HS, Some_Col, College)

# correlation matrix
Earnings_and_Height_interest <- Earnings_and_Height[,c("earnings", "height", "LT_HS", "HS", "Some_Col")]

cor(Earnings_and_Height_interest)
```

```
##              earnings      height       LT_HS          HS    Some_Col
## earnings   1.00000000  0.10428470 -0.22769147 -0.18970620  0.02474663
## height     0.10428470  1.00000000 -0.06625784 -0.04820749  0.01787849
## LT_HS     -0.22769147 -0.06625784  1.00000000 -0.25273987 -0.19240154
```

```
## HS       -0.18970620 -0.04820749 -0.25273987  1.00000000 -0.43467287
## Some_Col  0.02474663  0.01787849 -0.19240154 -0.43467287  1.00000000
```

```r
# data # women
Earnings_and_Height0 <- Earnings_and_Height[Earnings_and_Height$sex == 0,]

# Model 1
E72_M1 <- lm(formula = earnings ~ height,
             data = Earnings_and_Height0)

# Mdoel 2
E72_M2 <- lm(formula = earnings ~ height + LT_HS + HS + Some_Col,
             data = Earnings_and_Height0)

# Adjust standard errors
cov1 <- vcovHC(E72_M1, type = "HC1")
rb_se1 <- sqrt(diag(cov1))

cov2 <- vcovHC(E72_M2, type = "HC1")
rb_se2 <- sqrt(diag(cov2))

# output table
stargazer(E72_M1, E72_M2,
          type = "text",
          column.labels = c("Women", "Women"),
          se = list(rb_se1, rb_se2),
          digits = 4)
```

```
##
## ==========================================================================
##                                     Dependent variable:
##                     ------------------------------------------------------
##                                           earnings
##                               Women                      Women
##                                (1)                        (2)
## --------------------------------------------------------------------------
## height                      511.2222***                135.1421
##                             (97.5846)                  (92.3164)
##
## LT_HS                                               -31,857.8100***
##                                                        (834.9586)
##
## HS                                                  -20,417.8900***
##                                                        (637.8055)
##
## Some_Col                                            -12,649.0700***
##                                                        (716.5866)
##
## Constant                   12,650.8600**             50,749.5200***
##                            (6,299.1510)              (6,003.8190)
##
## --------------------------------------------------------------------------
## Observations                  9,974                     9,974
## R2                            0.0027                    0.1382
## Adjusted R2                   0.0026                    0.1378
```

```
## Residual Std. Error  26,800.9000 (df = 9972)   24,917.3800 (df = 9969)
## F Statistic         26.7214*** (df = 1; 9972) 399.6163*** (df = 4; 9969)
## =========================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01
```

i. The estimated coefficient on height falls by approximately 73.5649% when the education variables are added as control variables in the regression. This result coincides with our expectation.

```r
(E72_M2$coefficients[2] - E72_M1$coefficients[2]) / E72_M1$coefficients[2]
```

```
##    height
## -0.735649
```

ii. *College* is perfectly collinear with other education regressors and the constant regressor.

iii.

- **Prepare**

  $H_0 : \beta_{LT\_HS} = \beta_{HS} = \beta_{Some\_Col} = 0$ v.s. H_1 : not H_0

  Let the significance level be 5%.

- **Calculate**

```r
linearHypothesis(E72_M2,
                 c("LT_HS = 0", "HS = 0", "Some_Col = 0"),
                 test = "F")
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
9.97e+03	7.16e+12				
9.97e+03	6.19e+12	3	9.73e+11	523	2.13e-315

- **Conclusion**

  Because $p - value < \alpha = 5\%$, we reject H_0. There is significant evidence that educational differences exist.

iv.

- Other things being equal, workers with less than a high school education on average earn $31,857.81 less per year than a college graduate on average.
- Other things being equal, workers with a high school education on average earns $20,417.89 less per year than a college graduate on average.
- Other things being equal, workers with a some college on average earns $12,649.07 less per year than a college graduate on average.

c. Repeat (b), using data for men.

Solution

```r
# data # men
Earnings_and_Height1 <- Earnings_and_Height[Earnings_and_Height$sex == 1,]

# Model 3
E72_M3 <- lm(formula = earnings ~ height,
             data = Earnings_and_Height1)
```

```r
# Mdoel 4
E72_M4 <- lm(formula = earnings ~ height + LT_HS + HS + Some_Col,
             data = Earnings_and_Height1)

# Adjust standard errors
cov3 <- vcovHC(E72_M3, type = "HC1")
rb_se3 <- sqrt(diag(cov3))

cov4 <- vcovHC(E72_M4, type = "HC1")
rb_se4 <- sqrt(diag(cov4))

# output table
stargazer(E72_M3, E72_M4,
          type = "text",
          column.labels = c("Men", "Men"),
          se = list(rb_se3, rb_se4),
          digits = 4)
```

```
##
## ============================================================================
##                                      Dependent variable:
##                       -----------------------------------------------------
##                                           earnings
##                                 Men                         Men
##                                 (1)                         (2)
## ----------------------------------------------------------------------------
## height                    1,306.8600***               744.6809***
##                             (98.8569)                 (92.2615)
##
## LT_HS                                               -31,400.4900***
##                                                       (869.6952)
##
## HS                                                  -20,345.8500***
##                                                       (701.6438)
##
## Some_Col                                            -12,610.9200***
##                                                       (797.8023)
##
## Constant                 -43,130.3400***              9,862.7400
##                            (6,925.0110)              (6,541.3220)
##
## ----------------------------------------------------------------------------
## Observations                  7,896                       7,896
## R2                            0.0209                      0.1658
## Adjusted R2                   0.0207                      0.1654
## Residual Std. Error  26,671.2900 (df = 7894)    24,623.2200 (df = 7891)
## F Statistic         168.2010*** (df = 1; 7894) 392.0380*** (df = 4; 7891)
## ============================================================================
## Note:                                         *p<0.1; **p<0.05; ***p<0.01
```

i. The estimated coefficient on height falls by approximately 43.0175% when the education variables are added as control variables in the regression. This result coincides with our expectation.

```r
(E72_M4$coefficients[2] - E72_M3$coefficients[2]) / E72_M3$coefficients[2]
```

```
##     height
## -0.4301754
```

ii. *College* is perfectly collinear with other education regressors and the constant regressor.

iii.

- **Prepare**

  $H_0 : \beta_{LT\_HS} = \beta_{HS} = \beta_{Some\_Col} = 0$ v.s. H_1 : not H_0

  Let the significance level be 5%.

- **Calculate**

```r
linearHypothesis(E72_M4,
                 c("LT_HS = 0", "HS = 0", "Some_Col = 0"),
                 test = "F")
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
7.89e+03	5.62e+12				
7.89e+03	4.78e+12	3	8.31e+11	457	9.39e-274

- **Conclusion**

  Because $p - value < \alpha = 5\%$, we reject H_0. There is significant evidence that educational differences exist.

iv.

- Other things being equal, workers with less than a high school education on average earn $31,400.49 less per year than a college graduate on average.
- Other things being equal, workers with a high school education on average earns $20,345.85 less per year than a college graduate on average.
- Other things being equal, workers with a some college on average earns $12,610.92 less per year than a college graduate on average.