UN3412 Introduction to Econometrics

Recitation 5

Sean Hyland

Columbia University

Spring 2019

Outline

1. Course Evaluations and Beauty, continued
2. S&W Empirical Exercise 6.2
3. S&W Exercise 7.1
4. SW Exercise 7.4
5. S&W Additional Empirical Exercise 7.1

Outline

1. Course Evaluations and Beauty, continued
2. S&W Empirical Exercise 6.2
3. S&W Exercise 7.1
4. SW Exercise 7.4
5. S&W Additional Empirical Exercise 7.1

Course Evaluations and Beauty, continued (i)

Using the data set `TeachingRatings.dta` carry out the following exercises.

(a) Regress *Course_eval* on *Beauty* and *Female*, and test the hypothesis that all population coefficients are jointly significant at 5% significance level.

(b) Regress *Course_eval* on *Beauty*, *Female*, *Minority* and *Age*, test the hypothesis that all population coefficients are jointly significant at 5% significance level.

(c) Now test if *Minority* and *Age* are jointly significant at 1% significance level using the results from part (a) and part (b).

(d) Consider the various control variables in the data set. Which do you think should be included in the regression? Using a regression table, examine the effect of *Beauty* on *Course_eval*.

Outline

S&W Empirical Exercise 6.2 (i)

Using the data set `Growth`, described at `https://media.pearsoncmg.com/ph/bp/bp_stock_econometrics_4_cw/content/datapages/data/Growth_Description.pdf`, but excluding the data for Malta, carry out the following exercises.

(a) Construct a table that shows the sample mean, standard deviation, and minimum and maximum values for the series *Growth*, *TradeShare*, *YearsSchool*, *Oil*, *Rev_Coups*, *Assassinations*, and *RGDP60*. Include the appropriate units for all entries.

(b) Run a regression of *Growth* on *TradeShare*, *YearsSchool*, *Rev_Coups*, *Assassinations*, and *RGDP60*. What is the value of the coefficient on *Rev_Coups*? Interpret the value of this coefficient. Is it large or small in a real-world sense?

(c) Use the regression to predict the average annual growth rate for a country that has average values for all regressors.

(d) Repeat (c), but now assume that the country's value for *TradeShare* is one standard deviation above the mean.

(e) Why is *Oil* omitted from the regression? What would happen if it were included?

Outline

S&W Exercise 7.1 (i)

Add * (5%) and ** (1%) significance stars to the following table[1], to indicate the statistical significance of the coefficients.

[1] Note: I will use the corresponding table from the Updated 3rd Edition of the Textbook, whereas the problem set uses the table from the 3rd Edition.

Table 1: Results of Regressions of Average Hourly Earnings on 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
	(0.23)	(0.22)	(0.22)
Female (X_2)	-3.85	-3.81	-3.80
	(0.23)	(0.22)	(0.22)
Age (X_3)		0.51	0.52
		(0.04)	(0.04)
Northeast (X_4)			0.18
			(0.36)
Midwest (X_5)			-1.23
			(0.31)
South (X_6)			-0.43
			(0.30)
Intercept	17.02	1.87	2.05
	(0.17)	(1.18)	(1.18)
F-stat for regional effects=0			7.38
SER	9.79	9.68	9.67
R^2	0.162	0.180	0.182
n	7440	7440	7440

Note: Standard errors appear in parentheses.

S&W Exercise 7.1 (ii)

Add * (5%) and ** (1%) significance stars to the following table[1], to indicate the statistical significance of the coefficients.

Note, this requires the following assignment rule:

- If $|t| > 2.576 \Rightarrow$ two significance stars,
- Else if $2.576 > |t| > 1.96 \Rightarrow$ one significance stars,
- Else if $1.96 > |t| \Rightarrow$ no significance stars,

[1] Note: I will use the corresponding table from the Updated 3rd Edition of the Textbook, whereas the problem set uses the table from the 3rd Edition.

Table 1: Results of Regressions of Average Hourly Earnings on 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**
	(0.23)	(0.22)	(0.22)
Female (X_2)	-3.85**	-3.81**	-3.80**
	(0.23)	(0.22)	(0.22)
Age (X_3)		0.51**	0.52**
		(0.04)	(0.04)
Northeast (X_4)			0.18
			(0.36)
Midwest (X_5)			-1.23**
			(0.31)
South (X_6)			-0.43
			(0.30)
Intercept	17.02**	1.87	2.05
	(0.17)	(1.18)	(1.18)
F-stat for regional effects=0			7.38
SER	9.79	9.68	9.67
R^2	0.162	0.180	0.182
n	7440	7440	7440

Note: Standard errors appear in parentheses.

Outline

Q2: SW Exercise 7.4

Using Column (3) of Table 1:

(a) Do there appear to be important regional differences? Use an appropriate hypothesis test to explain your answer.

Table 1: Results of Regressions of Average Hourly Earnings on 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**
	(0.23)	(0.22)	(0.22)
Female (X_2)	-3.85**	-3.81**	-3.80**
	(0.23)	(0.22)	(0.22)
Age (X_3)		0.51**	0.52**
		(0.04)	(0.04)
Northeast (X_4)			0.18
			(0.36)
Midwest (X_5)			-1.23**
			(0.31)
South (X_6)			-0.43
			(0.30)
Intercept	17.02**	1.87	2.05
	(0.17)	(1.18)	(1.18)
F-stat for regional effects=0			7.38
SER	9.79	9.68	9.67
R^2	0.162	0.180	0.182
n	7440	7440	7440

Note: Standard errors appear in parentheses.

Q2: SW Exercise 7.4

Using Column (3) of Table 1:

(a) Do there appear to be important regional differences? Use an appropriate hypothesis test to explain your answer.
The 1% critical value of the F(3, ∞) distribution is 3.78, less than the F-statistic associated with the null hypothesis that all regional effects are zero (7.38), thus we reject the null hypothesis.

Q2: SW Exercise 7.4

Using Column (3) of Table 1:

(b) Juanita is a 28-year-old female college grad. from the South.
Molly is a 28-year-old female college grad. from the West.
Jennifer is a 28-year-old female college grad. from the Midwest.

(i) Construct a 95% CI for the difference in expected earnings between Juanita and Molly.

(ii) Explain how you would construct a 95% CI for the difference between Juanita and Jennifer?

Table 1: Results of Regressions of Average Hourly Earnings on 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**
	(0.23)	(0.22)	(0.22)
Female (X_2)	-3.85**	-3.81**	-3.80**
	(0.23)	(0.22)	(0.22)
Age (X_3)		0.51**	0.52**
		(0.04)	(0.04)
Northeast (X_4)			0.18
			(0.36)
Midwest (X_5)			-1.23**
			(0.31)
South (X_6)			-0.43
			(0.30)
Intercept	17.02**	1.87	2.05
	(0.17)	(1.18)	(1.18)
F-stat for regional effects=0			7.38
SER	9.79	9.68	9.67
R^2	0.162	0.180	0.182
n	7440	7440	7440

Note: Standard errors appear in parentheses.

Q2: SW Exercise 7.4

Using Column (3) of Table 1:

(b) Juanita is a 28-year-old female college grad. from the South.
Molly is a 28-year-old female college grad. from the West.
Jennifer is a 28-year-old female college grad. from the Midwest.

(i) Construct a 95% CI for the difference in expected earnings between Juanita and Molly.
Since Molly is from the omitted category, the difference between the expected earnings for Juanita and Molly is simply the coefficient on the indicator variable for South, and the 95% confidence interval is given by:
$-0.43 \pm (1.96 \times 0.30) = [-1.02; 0.16]$

(ii) Explain how you would construct a 95% CI for the difference between Juanita and Jennifer?

Table 1: Results of Regressions of Average Hourly Earnings on 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**
	(0.23)	(0.22)	(0.22)
Female (X_2)	-3.85**	-3.81**	-3.80**
	(0.23)	(0.22)	(0.22)
Age (X_3)		0.51**	0.52**
		(0.04)	(0.04)
Northeast (X_4)			0.18
			(0.36)
Midwest (X_5)			-1.23**
			(0.31)
South (X_6)			-0.43
			(0.30)
Intercept	17.02**	1.87	2.05
	(0.17)	(1.18)	(1.18)
F-stat for regional effects=0			7.38
SER	9.79	9.68	9.67
R^2	0.162	0.180	0.182
n	7440	7440	7440

Note: Standard errors appear in parentheses.

Q2: SW Exercise 7.4

Using Column (3) of Table 1:

(b) Juanita is a 28-year-old female college grad. from the South.
Molly is a 28-year-old female college grad. from the West.
Jennifer is a 28-year-old female college grad. from the Midwest.

(i) Construct a 95% CI for the difference in expected earnings between Juanita and Molly.

Since Molly is from the omitted category, the difference between the expected earnings for Juanita and Molly is simply the coefficient on the indicator variable for South, and the 95% confidence interval is given by:
$-0.43 \pm (1.96 \times 0.30) = [-1.02; 0.16]$

(ii) Explain how you would construct a 95% CI for the difference between Juanita and Jennifer?

Could use the general methods of Section 7.3, or... run a regression on the same data, this time omitting Midwest and including West. Then the coefficient on South, and its standard error, are informative of the difference between the South and the Midwest.

Outline

S&W Additional Empirical Exercise 7.1 (i)

Use the data set `CPS12.dta` described in AEE 4.1 to answer the following:

(a) Run a regression of average hourly earnings (*AHE*) on age (*Age*).
- What is the estimated intercept?
- What is the estimated slope?

(b) Run a regression of *AHE* on *Age*, gender (*Female*), and education (*Bachelor*).
- What is the estimated effect of *Age* on earnings?
- Construct a 95% confidence interval for the coefficient on *Age* in the regression.

(c) Are the results from the regression in (b) substantively different from the results in (a) regarding the effects of *Age* and *AHE*? Does the regression in (a) seem to suffer from omitted variable bias?

(d) Bob is a 26-year-old male worker with a high school diploma. Predict Bob's earnings using the estimated regression in (b).
Alexis is a 30-year-old female worker with a college degree. Predict Alexis's earnings using the regression.

S&W Additional Empirical Exercise 7.1 (ii)

(e) Compare the fit of the regression in (a) and (b) using the regression standard errors, R^2 and $\bar{R}^2$. Why are the R^2 and $\bar{R}^2$ so similar in regression (b)?

(f) Are gender and education determinants of earnings?

- Test the null hypothesis that *Female* can be deleted from the regression.
- Test the null hypothesis that *Bachelor* can be deleted from the regression.
- Test the null hypothesis that both *Female* and *Bachelor* can be deleted from the regression.

(g) A regression will suffer from omitted variable bias when two conditions hold.

- What are these two conditions?
- Do these conditions seem to hold here?