Answers for Homework #2

Zheng Tian

1 Theoretical Exercises

4.2 The estimated regression equation is

$$\widehat{Weight} = -99.42 + 3.94 \times Height,\ R^2 = 0.81,\ SER = 10.2$$

a. Substituting $Height = (70, 65, 74)$ inches into the equation, the predicted weights are $(176.39, 156.69, 192.15)$ pounds, respectively.

b. $\Delta\widehat{Weight} = 3.94 \times \Delta Height = 3.94 \times 1.5 = 5.91$ inches.

c. Let's consider this problem from a general case. Suppose the original estimated regression model is

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i,\ i = 1, \ldots, n$$

Now we have new data with different units such that $x_i = aX_i$, $y_i = bY_i$. It is easy to see that

$$\bar{x} = a\bar{X},\ \bar{y} = b\bar{Y}, \sum_i (x_i - \bar{x}) = a \sum_i (X_i - \bar{X}),\ \text{ and } \sum_i (y_i - \bar{y}) = b \sum_i (Y_i - \bar{Y})$$

Let $\tilde{\beta}_0$ and $\tilde{\beta}_1$ be the estimated coefficients and $\tilde{u}_i$ be the residuals in the new regression equation as follows,

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_i + \tilde{u}_i$$

Then, we have

$$\tilde{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{ab \sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{a^2 \sum_i (X_i - \bar{X})} = \frac{b}{a}\hat{\beta}_1$$

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x} = b\bar{Y} - \frac{b}{a}\hat{\beta}_1 (a\bar{X}) = b\hat{\beta}_0$$

$$\tilde{u}_i = y_i - \hat{y}_i = b(Y_i - \hat{Y}_i) = b\hat{u}_i$$

$$\tilde{R}^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum_i \tilde{u}_i^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{b^2 \sum_i \hat{u}_i^2}{b^2 \sum_i (Y_i - \bar{Y})^2} = R^2$$

$$\widetilde{SER} = \sqrt{\frac{1}{n-2} \sum_i \tilde{u}_i^2} = \sqrt{\frac{b^2}{n-2} \sum_i \hat{u}_i^2} = b\, SER$$

Now let's go back to the specific question. We know that 1 inch = 2.54 cm and 1 pound = 0.4536 kg so that $Weight_{new} = 0.4536 \times Weight$ and $Height_{new} =$

$2.54 \times Height$. Thus, using the results above, we obtain

$$\tilde{\beta}_1 = (0.4536/2.54) \times 3.94 = 0.7036$$
$$\tilde{\beta}_0 = -99.41 \times 0.4536 = -45.0924$$
$$\tilde{R}^2 = 0.08,\ \widetilde{SER} = 0.4536 \times 10.2 = 4.6267$$

4.3 The estimated regression equation is

$$\widehat{AWE} = 696.7 + 9.6 \times Age,\ R^2 = 0.023,\ SER = 624.1$$

a. The coefficient 9.6 shows the marginal effect of Age on AWE; that is, AWE is expected to increase by 9.6 for each additional year of age. 696.7 is the intercept of the regression line. It determines the overall level of the line.

b. SER is in the same units as the dependent variable (Y, or AWE in this example). Thus SER is measured in dollars per week.

c. R^2 is unit free.

d. Plugging 25 and 45 into the regression equation,

- $696.7 + 9.6 \times 25 = 936.7$
- $696.7 + 9.6 \times 45 = 1128.7$

e. No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.

f. No. The distribution of earning is positively skewed and has kurtosis larger than the normal.

g $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$. Thus, the sample mean of AWE is $696.7 + 9.6 \times 41.6 = 1096.06$.

4.5 a. u_i represents factors other than time that influence the student's performance on the exam including amount of time studying, aptitude for the material, and so forth. Some students will have studied more than average, other less; some students will have higher than average aptitude for the subject, others lower, and so forth.

b. Because of random assignment u_i is independent of X_i. Since u_i represents deviations from average $E(u_i) = 0$. Because u and X are independent $E(u_i|X_i) = E(u_i) = 0$.

c. Assumption #2 is satisfied if this year's class is typical of other classes, that is, students in this year's class can be viewed as random draws from the population of students that enroll in the class. Assumption #3 is satisfied because both X and Y are bounded.

d.
- 70.6 for 95 minutes; 77.8 for 120 minutes; 85.0 for 150 minutes
- 2.4 for 10 more minutes.

4.10 a. Assumption #1 is satisfied since whatever value X takes we always have $E(u_i) = 0$. Assumption #2 is satisfied because (u_i, X_i) is i.i.d and Y_i is a function of X_i and u_i. X_i is bounded and so has finite fourth moment; the fourth moment is non-zero because $\Pr(X_i = 0)$ and $\Pr(X_i = 1)$ are both non-zero so that X_i has finite, non-zero kurtosis. Following calculation like those exercise 2.13, u_i also has non-zero finite fourth moment.

b. $\text{var}(X_i) = 0.2 \times (1-0.2) = 0.16$ and $\mu_X = 0.2$. Also,

$$
\begin{aligned}
&\text{var}\left((X_i - \mu_X)u_i\right) = E\left((X_i-\mu_X)u_i\right)^2 = E\left[E\left((X_i-\mu_X)u_i\right)^2 |X\right] \\
&= E\left[\left((X_i-\mu_X)u_i\right)^2 |X_i = 0\right] \cdot \Pr(X_i = 0) + E\left[\left((X_i-\mu_X)u_i\right)^2 |X_i = 1\right] \cdot \Pr(X_i = 1) \\
&= E((0-0.2)^2 u_i^2) \times 0.8 + E((1-0.2)^2 u_i^2) \times 0.2 \\
&= 0.2^2 \times 1 \times 0.8 + 0.8^2 \times 4 \times 0.2 \\
&= 0.544
\end{aligned}
$$

Therefore,

$$
\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{\text{var}\left((X_i-\mu_X)u_i\right)}{[\text{var}(X_i)]^2} = \frac{1}{n}\frac{0.544}{0.16^2} = \frac{1}{n}21.25
$$

4.12 a. Write

$$
\begin{aligned}
ESS &= \sum_i (\hat{Y}_i - \bar{Y})^2 = \sum_i (\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{Y})^2 = \sum_i \left[\hat{\beta}_1 (X_i - \bar{X})\right]^2 \\
&= \hat{\beta}_1^2 \sum_i (X_i - \bar{X})^2 = \frac{\left[\sum_i (X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum_i (X_i - \bar{X})^2}
\end{aligned}
$$

This implies

$$
\begin{aligned}
R^2 = \frac{ESS}{TSS} &= \frac{\left[\sum_i (X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2} \\
&= \left[\frac{\frac{1}{n-1}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\left(\frac{1}{n-1}\sum_i (X_i - \bar{X})^2\right)^{1/2}\left(\frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2\right)^{1/2}}\right]^2 \\
&= \left[\frac{s_{XY}}{s_X s_Y}\right]^2 = r_{XY}^2
\end{aligned}
$$

b. This follows from part (a) because $r_{XY} = r_{YX}$.

c. $r_{XY}\frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2} = \frac{\frac{1}{n-1}\sum_i (X_i-\bar{X})(Y_i-\bar{Y})}{\frac{1}{n-1}\sum_i (X_i-\bar{X})^2} = \frac{\sum_i (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_i (X_i-\bar{X})^2} = \hat{\beta}_1$

2 Empirical Exercise

This file include answers and R codes for completing Empirical Exercise 4.2 in Introduction to Econometrics (3rd edition) by Stock and Watson.

Reading the Data

The first step is to read the data file into R. The data files for this problem are `TeachingRatings.dta` and `TeachingRatings.xls`, accompanied by a descriptive file `TeachingRatings_Description.pdf`.

- Read the STATA file

  ```
  library(foreign)
  teachingdata <- read.dta("TeachingRatings.dta")
  ```

- Upon reading the data, we can take a glimpse on the data.
  - Use `head` or `tail` to look at the first or last few observations

    ```
    head(teachingdata)
    ```

Summary Statistics

We get the summary statistics of the variables used in the analysis, which is `course_eval` and `beauty`

```
df <- teachingdata[c("course_eval", "beauty")]
sumdf <- summary(df); sumdf
```

```
 course_eval         beauty
Min.   :2.100   Min.   :-1.45049
1st Qu.:3.600   1st Qu.:-0.65627
Median :4.000   Median :-0.06801
Mean   :3.998   Mean   : 0.00000
3rd Qu.:4.400   3rd Qu.: 0.54560
Max.   :5.000   Max.   : 1.97002
```

We can create a table that looks professional using `stargazer()`.

```
library(stargazer)
stargazer(df, type = "latex",
  title = "Summary Statistics", label = "tab:sum-stats")
```

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
course_eval	463	3.998	0.555	2.100	5.000
beauty	463	0.00000	0.789	−1.450	1.970

Scatterplot

We can make scatterplot using the `plot` function.

```
teaching.formula <- course_eval ~ beauty
plot(teaching.formula, data = teachingdata,
   main = "The Scatterplot of Course Evaluation on Professor's Beauty",
   xlab="Beauty", ylab = "Course evaluation", col = "blue")
```

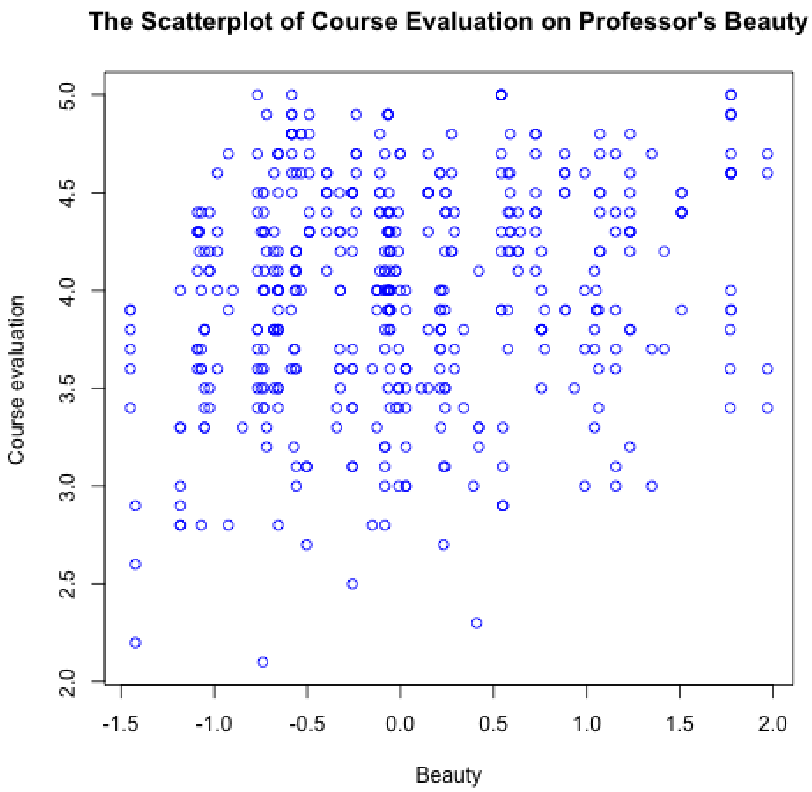

Figure 1: The scatterplot of course evaulation on professors' beauty

Regression

Now let's estimate the regression model. The results is reported in Table 2

```
# run a regression of course evaluation on professor's beauty
teaching.ols <- lm(teaching.formula, data = teachingdata)

# create the latex table
stargazer(teaching.ols,
  covariate.labels = c("Beauty"),
  dep.var.labels = c("Course Evaluations"),
  title = "The OLS Estimation of the Regression of Course Evaluation on Beauty",
  label = "tab:ols-1", single.row = TRUE, omit.stat = c("adj.rsq", "f")
)
```

Table 2: The OLS Estimation of the Regression of Course Evaluation on Beauty

	Dependent variable:
	Course Evaluations
Beauty	0.133*** (0.032)
Constant	3.998*** (0.025)
Observations	463
R^2	0.036
Residual Std. Error	0.545 (df = 461)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answers to the questions

a. The scatterplot is Figure 1. There appears to be a weak positive relationship between course evaluation and the beauty index.

b. The estimation results are reported in Table 2.

```
beauty.watson <- mean(teachingdata$beauty)
beauty.stock <- mean(teachingdata$beauty) + sd(teachingdata$beauty)
ave.courseval <- mean(teachingdata$course_eval)

# do prediction step by step
b0 <- teaching.ols$coef[1]
b1 <- teaching.ols$coef[2]
courseval.predict <- b0 + b1 * c(beauty.watson, beauty.stock)
names(courseval.predict) <- c("waston", "stock")
```

The slope is `0.133` and the intercept is `3.998`. The sample mean of course evaluation is `3.998`, which coincides with the slope because the sample mean of *Beauty* is `0`.

c. The beauty indices for Professors Stock and Watson are `0.7886` (one standard deviation) and `0` (sample average). Thus, the predicted course evaluations for Professors Stock and Watson are `4.1032` and `3.9983`, respectively.

```
beauty.sd <- sd(teachingdata$beauty)
courseval.sd <- sd(teachingdata$course_eval)
delta.courseval <- b1 * beauty.sd
```

d. The standard deviation of course evaluation is `0.5549`, and the standard deviation of beauty is `0.7886`. A one-standard-deviation increase in beauty is expected to increase course evaluation by `0.1049`, or `0.19` of standard deviation of course evaluations. The effect is small.

```
rsq <- summary(teaching.ols)$r.squared
```

e. The regression R^2 is `0.0357`, so that *Beauty* explains only `3.6` percent of the variance in course evaluations.