Contents

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

1 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)
    ```

2 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

3 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")
```

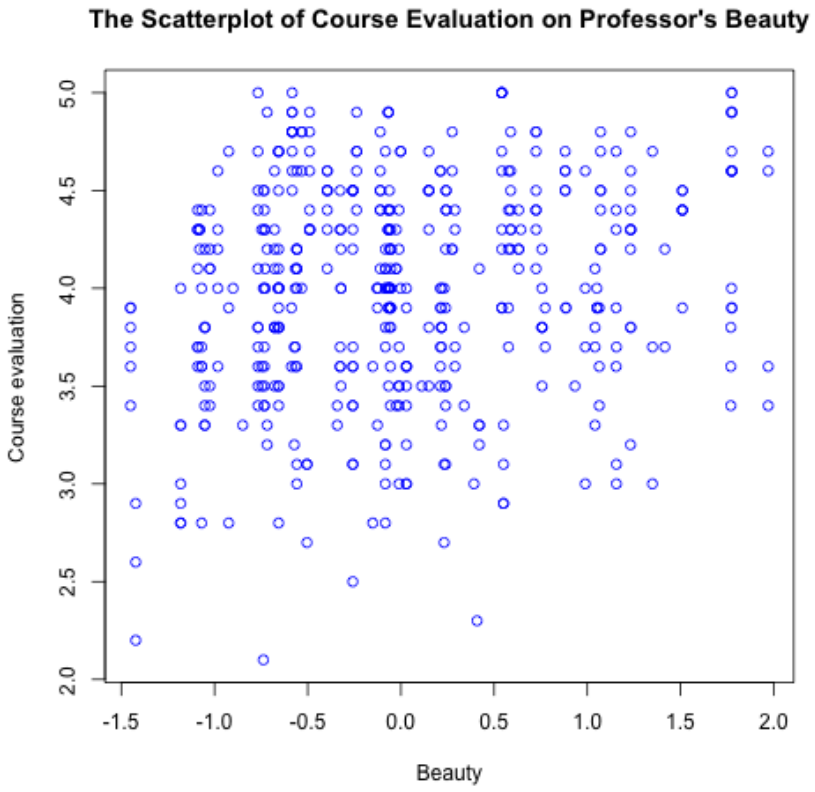

Figure 1: The scatterplot of course evaulation on professors' beauty

4 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

5 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.