Homework2 Empirical Exercise

Qianhui Yang

3/27/2017

E4.3

a.

Construct a scatterplot of average course evaluations (Course_Eval) on the professor's beauty (Beauty). Does there appear to be a relationship between the variables?

```
   library(foreign)
classdata <- read.dta("C:/econometrics/teachingRate/TeachingRatings.dta")
```

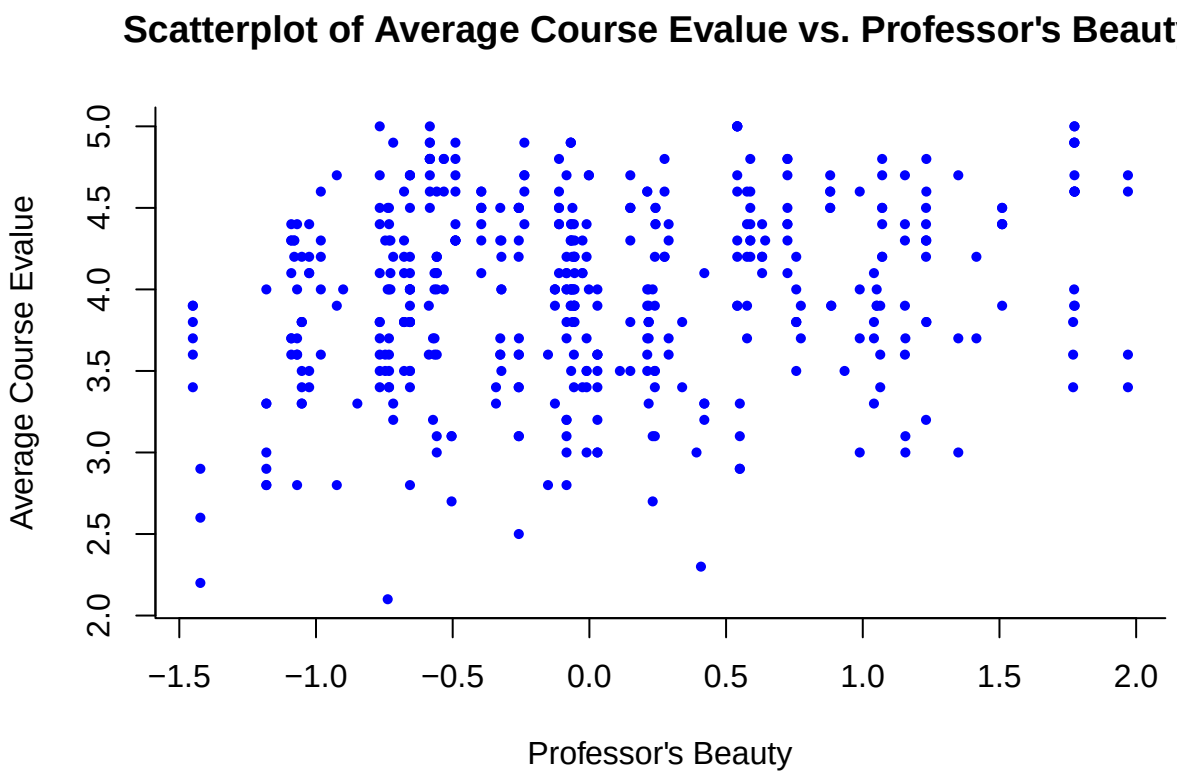
```
  head(classdata[c("beauty", "course_eval")])
```

```
##       beauty course_eval
## 1  0.2899157         4.3
## 2 -0.7377322         4.5
## 3 -0.5719836         3.7
## 4 -0.6779634         4.3
## 5  1.5097942         4.4
## 6  0.5885687         4.2
```

```
df <- classdata[c("beauty", "course_eval")]
summary(df)
```

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

```
  # generate a scatterplot
  plot(df$beauty, df$course_eval, col = "blue", pch =16, cex = 0.7, bty = "l",
       main = "Scatterplot of Average Course Evalue vs. Professor's Beauty",
       xlab = "Professor's Beauty", ylab = "Average Course Evalue")
```



Yes,there appear to be a positive relationship between the varibles.

b.

Run a regression of average course evaluations (Course_Eval) on the professor's beauty (Beauty). What is the estimated intercept? What is the estimated slope? Explain why the estimated intercept is equal to the sample mean of Course_Eval. (Hint: What is the sample mean of Beauty?)

$$CourseEvalue_i = \beta_0 + \beta_1 Beauty_i + u_i$$

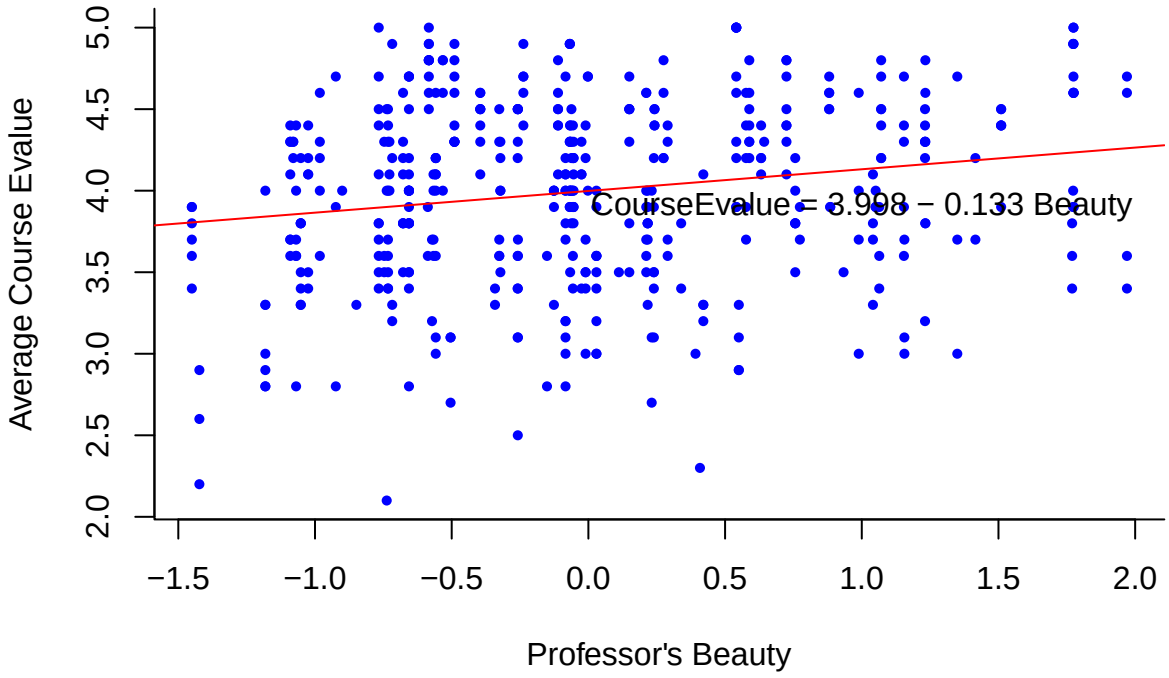
```
mod1 <- lm(course_eval ~ beauty, data = df)
summary(mod1)
```

```
##
## Call:
## lm(formula = course_eval ~ beauty, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.80015 -0.36304  0.07254  0.40207  1.10373
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.99827    0.02535 157.727  < 2e-16 ***
```

```
## beauty       0.13300    0.03218   4.133 4.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5455 on 461 degrees of freedom
## Multiple R-squared:  0.03574,    Adjusted R-squared:  0.03364
## F-statistic: 17.08 on 1 and 461 DF,  p-value: 4.247e-05
```

$$\widehat{CourseEvalue} = 3.998 - 0.133 \times Beauty$$

```
plot(df$beauty, df$course_eval,
     col = "blue", pch =16, cex = 0.7, bty = "l",
     xlab = "Professor's Beauty", ylab = "Average Course Evalue")
# add a straight line with an intercept a and slop b
abline(coef(mod1)[1], coef(mod1)[2], col="red")
# add a text on the plot
text(1, 3.9, "CourseEvalue = 3.998 - 0.133 Beauty",
     cex.lab = 0.03, font.lab = 7)
```



The estimated intercept is 3.998, and the estimated slope is 0.133. The estimated intercept is equal to the sample mean of Course_Evalue, because it is actually equal to $\bar{CourseEvalue} - \bar{Beauty} * \widehat{solpe}$ and $\bar{Beauty} = 0$.

c.

Professor Watson has an average value of Beauty, while Professor Stock's value of Beauty is one standard deviation above the average. Predict Professor Stock's and Professor Watson's course evaluations.

```
  # Replicate the summary statistics in Table 1.1
  summary1.1 <- function(df) {
    ave <- sapply(df, mean)
    std <- sapply(df, sd)
    perctile <- sapply(df, function(x)
    quantile(x, probs = c(0.1, 0.25, 0.4, 0.5, 0.6, 0.75, 0.9)))
    return(rbind(ave, std, perctile))
  }
  library(xtable)
  sumtab <- xtable(t(summary1.1(df)))
```

```
  # print as a latex table
print(sumtab, type = "latex")
```

% latex table generated in R 3.3.2 by xtable 1.8-2 package % Tue Apr 04 10:57:46 2017

	ave	std	10%	25%	40%	50%	60%	75%	90%
beauty	0.00	0.79	-0.98	-0.66	-0.26	-0.07	0.03	0.55	1.15
course_eval	4.00	0.55	3.30	3.60	3.90	4.00	4.20	4.40	4.70

Professor Watson: Course_Evalue=3.998-0.133*0=3.998

Professor Stock: Course_EValue=3.998-0.133*0.789=3.893

d.

Comment on the size of the regression's slope. Is the estimated effect of Beauty on Course_Eval large or small? Explain what you mean by "large" and "small".

The slope is equal to 0.133,which is quite small,so the estimated effect of Beauty on Course_Evalue is small.

e.

Does Beauty explain a large fraction of the variance in evaluations across courses? Explain.

As $R^2 = 0.0357$, the fraction of the sample variance of Course_Evalue explained by Beauty is 3.57%. So it doesn't explain a large fraction.