HW4

Zihan Lin

2023-02-06

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

1. Stock and Watson, exercise 4.3

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following: $A\hat{W}E = 696.7 + 9.6 * Age, R^2 = 0.023, SER = 624.1$.

a. Explain what the coefficient values 696.7 and 9.6 mean.

696.7 in the equation represents the value β_0 or the y intercept. Which can be described as the minimum average earning. 9.6 is β_1, or the estimator of the equation. Which can be described as the rate of the growth for AWE as the we increase the age.

b.The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)

Dollars

c. The regression R2 is 0.023. What are the units of measurement for the R2? (Dollars? Years? Or is R2 unit free?)

Unit free

d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?

$A\hat{W}E_{25} = 696.7 + 9.6 * 25 = 936.7$ dollars
$A\hat{W}E_{45} = 696.7 + 9.6 * 45 = 1128.7$ dollars

e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?

No, it will never give a reliable prediction for a 99-year-old worker, since the regression line is composed base on random sample of college-educated full-time workers aged 25–65. 99 is not inside this sample.

f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)

No, the graph will skewed to the right, since there is more college-educated full-time workers are younger, the mean will shifted to the left. Therefore, it is not normal.

g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample? (Hint: Review Key Concept 4.2.)

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Then we will have $A\hat{W}E = 696.7 + 9.6 * 41.6 = 1096.06$ dollars

2. Stock and Watson, exercise 4.4

Read the box "The 'Beta' of a Stock" in Section 4.2.

a. Suppose the value of β is greater than 1 for a particular stock. Show that the variance of $(R - R_f)$ for this stock is greater than the variance of $(R_m - R_t)$.

$\beta > 1$, $R - R_f = \beta(R_m - R_t)$, then $(R_m - R_t) < (R - R_f)$

b. Suppose the value of β is less than 1 for a particular stock. Is it possible that the variance of $(R - Rf)$ for this stock is greater than the variance of $(Rm - Rt)$? (Hint: Don't forget the regression error.)

$\beta < 1$, $R - R_f = \beta(R_m - R_f)$, $R - R_f = \beta(R_m - R_f) + u_i$. We have that $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + u_i$ where u_i is the error factor of the regression line. We do have that the return can be lower for $(Rm - Rt)$ than $(R_m - R_t)$.

c. In a given year, the rate of return on 3-month Treasury bills is 2.0% and the rate of return on a large diversified portfolio of stocks (the S&P 500) is 5.3%. For each company listed in the table in the box, use the estimated value of b to estimate the stock's expected rate of return.

```
beta <- c(0.1,0.6,0.7,1.0,1.1,1.3,1.7)
name <- c("Wal-Mart (discount retailer)","Coca-Cola (soft drinks)","Verizon (telecommunications)","Goog
df<- data.frame(beta,name)
df <- df%>% mutate("new_beta" = 0.053 + 0.02*(beta))
df
```

```
##   beta                            name new_beta
## 1  0.1    Wal-Mart (discount retailer)    0.055
## 2  0.6         Coca-Cola (soft drinks)    0.065
## 3  0.7    Verizon (telecommunications)    0.067
## 4  1.0 Google (information technology)    0.073
## 5  1.1   General Electric (industrial)    0.075
## 6  1.3               Boeing (aircraft)    0.079
## 7  1.7            Bank of America (bank)    0.087
```

3. Show that the R^2 from the regression of Y on X is the same as the R^2 of the regression of X on Y.

$R^2 = \frac{\Sigma_{i=1}^n(\hat{Y}_i - \bar{Y})^2}{\Sigma_{i=1}^n(Y_i - \bar{Y})^2}$
The regression R^2 is the fraction of the sample variance of Y explained by (or predicted by) X. Then if we have same sample variance of X and Y, then that the R^2 from the regression of Y on X is the same as the R^2 of the regression of X on Y.

4. Derive the least squares estimators when there is one regressor X. You many follow the outline given in Appendix 4.2, but make sure to fill in the details (i.e., show your work, including necessary algebra).

function 1: $\frac{d}{d\beta_0}\Sigma_{i=1}^n(Y_i - \beta_0 - \beta_1 X_i)^2 = -2\Sigma_{i=1}^n(Y_i - \beta_0 - \beta_1 X_i)$
function 2: $\frac{d}{d\beta_1}\Sigma_{i=1}^n(Y_i - \beta_0 - \beta_1 X_i)^2 = -2\Sigma_{i=1}^n(Y_i - \beta_0 - \beta_1 X_i)X_i$
To set the square minimize, we will have that function 1 and function 2 equal to 0. Or $\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1\bar{X} = 0$. For the statement above, we move the constant $\hat{\beta}_0$ to the other side of the equation, we can get the regression formula $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$

5. Stock and Watson, exercise 5.2 (note: you do not need to compute the p-value for part b)

Suppose a researcher, using wage data on 250 randomly selected male workers and 280 female workers, estimates the OLS regression
$A\hat{W}E = 12.52 + 2.12 * Male, R^2 = 0.06, SER = 4.2, SE(\beta_0) = 0.23, SE(\beta_1) = 0.36$
where Wage is measured in dollars per hour and Male is a binary variable that is equal to 1 if the person is a male and 0 if the person is a female. Define the wage gender gap as the difference in mean earnings between men and women.

a. What is the estimated gender gap?

$\hat{\beta}_1 = 2.12$

b. Is the estimated gender gap significantly different from 0? (Compute the p-value for testing the null hypothesis that there is no gender gap.)

$t = \frac{2.12-0}{0.36} = 5.889$, p-value = 0 There is significantly different from 0.

c. Construct a 95% confidence interval for the gender gap.

$\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1) = 2.12 \pm 1.96 \times 0.36.$
Confidence Interval: $\{1.4144, 2.8256\}$

d. In the sample, what is the mean wage of women? Of men?

$\mu_{men} = 12.52 + 2.12 * 1 = 14.64$
$\mu_{women} = 12.52 + 2.12 * 0 = 12.52$

e. Another researcher uses these same data but regresses Wages on Female, a variable that is equal to 1 if the person is female and 0 if the person a male. What are the regression estimates calculated from this regression?

$A\hat{W}E = 14.64 - 2.12 * Female, R^2 = 0.06, SER = 4.2$

6. Stock and Watson, exercise 5.5

In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to "regular" and "small" classes and given standardized tests at the end of the year. (Regular classes contained approxi- mately 24 students, and small classes contained approximately 15 students.) Suppose, in the population, the standardized tests have a mean score of 925 points and a standard deviation of 75 points. Let SmallClass denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise. A regression of TestScore on SmallClass yields
$Test\hat{S}core = 918.0 + 13.9 * SmallClass, R^2 = 0.01, SER = 74.6, SE(\beta_0) = 1.6, SE(\beta_1) = 2.5$

a. Do small classes improve test scores? By how much? Is the effect large? Explain.

Yes, small classes improve test scores by 13.9 points. This is a large effect base on $\hat{\beta}_0$

b. Is the estimated effect of class size on test scores statistically significant? Carry out a test at the 5% level.

$\frac{13.9-0}{2.5} = 5.56$, p-value $= 0$. It is statistically significant.

c. Construct a 99% confidence interval for the effect of SmallClass on TestScore.

$\hat{\beta}_1 \pm 2.576 \times SE(\hat{\beta}_1) = 13.9 \pm 2.576 \times 2.5.$
Confidence Interval: $\{7.46, 20.34\}$

d. Does least squares assumption 1 plausibly hold for this regression? Explain.

No its not. Base on the experiment, we can tell that there are many factors will affect the test score from a student. For instance, food they eat, their healthy condition, etc. Therefore, the effect from other factor u_i can't be determined as zero at this point.