Week 11 - Problems

Business Research Methods

Bhaswar Chakma

04 May 2021

Problem 1: CEO Salary

Install the package `wooldridge`

- data: `wooldridge::ceosal1`

The data set `ceosal1` contains information on CEOs.

- `salary`: salary in thousands of dollars.
- `roe`: average return on equity (roe) for the CEO's firm for the previous three years.

Q1

1. Sample size? Find average, minimum and maximum salary.

```
# find n; avg, min, max salary
ceosal1 %>%
  summarise(n = n(),
            avg_sal = mean(salary),
            min_sal = min(salary),
            max_sal = max(salary))
```

```
##     n avg_sal min_sal max_sal
## 1 209 1281.12     223   14822
```

Estimate the model

$$salary = \beta_0 + \beta_1 roe + u.$$

```
model_ceo <- lm(salary ~ roe, data = ceosal1)
model_ceo$coefficients
```

```
## (Intercept)         roe
##   963.19134    18.50119
```

$$\widehat{salary} = 963.191 + 18.501 roe$$

Q2

2. If `roe` is 0, what is the predicted salary?

$$\widehat{salary} = 963.191 + 18.501 roe$$

If the return on equity is zero, $roe = 0$, then the predicted salary is the intercept, 963.191, which equals \$963,191 because salary is measured in thousands.

Q3

$$\widehat{salary} = 963.191 + 18.501 roe$$

3. If `roe` increases by one percentage point, then the salary is predicted to change by ?

If the return on equity increases by one percentage point, $roe = 1$, then salary is predicted to change by about 18.5, or \$18,500.

Q4

4. If `roe` is 30, what is the predicted salary?

$$\widehat{salary} = 963.191 + 18.501(30) = 1518.221,$$

which is just over \$1.5 million.

Q5

5. *The firm's return on equity explains only about 1.3% of the variation in salaries for this sample of 209 CEOs. That means that 98.7% of the salary variations for these CEOs is left unexplained!* — True or False?

```
summary(model_ceo)$r.squared
```

```
## [1] 0.01318862
```

True!

Q6

6. *The first four CEOs have lower salaries than what we predicted from the OLS regression line. — True or False?*

```
broom::augment(model_ceo) %>% head(n = 4)
```

```
## # A tibble: 4 x 8
##    salary   roe .fitted .resid     .hat .sigma    .cooksd .s
##     <int> <dbl>   <dbl>  <dbl>    <dbl>  <dbl>      <dbl>
## 1    1095 14.1    1224.  -129. 0.00541  1370.  0.0000244
## 2    1001 10.9    1165.  -164. 0.00740  1370.  0.0000540
## 3    1122 23.5    1398.  -276. 0.00743  1370.  0.000154
## 4     578  5.90   1072.  -494. 0.0132   1369.  0.000888
```

True!

Q7

7. *The fifth CEO makes more than predicted from the OLS regression line.* — True or False?

```
broom::augment(model_ceo) %>% slice(5)
```

```
## # A tibble: 1 x 8
##   salary   roe .fitted .resid    .hat .sigma    .cooksd .s
##    <int> <dbl>   <dbl>  <dbl>   <dbl>  <dbl>      <dbl>
## 1   1368  13.8   1219.   149. 0.00554  1370. 0.0000335
```

True!

Q8

8. Rather than measuring `salary` in thousands of dollars, we want to measure it in dollars. Mutate a new variable `salardol` using this expression `salardol = 1000*salary`. Now estimate the following model:

$$salarydol = \beta_0 + \beta_1 roe + u.$$

```r
# mutate
ceosal1_new <- ceosal1 %>%
  mutate(salarydol = salary*1000)
# model
model_ceo2 <- lm(salarydol ~ roe, data = ceosal1_new)
```

```r
model_ceo$coefficients # salary in thousands of dollars
```

```
## (Intercept)         roe
##   963.19134    18.50119
```

```r
model_ceo2$coefficients # salary in dollars
```

```
## (Intercept)         roe
##   963191.34    18501.19
```

If the dependent variable is multiplied by the constant c – which means each value in the sample is multiplied by c – then the OLS intercept and slope estimates are also multiplied by c. — True!

Problem 2: College GPA

`wooldridge::gpa1`

- `colGPA`: college grade point average
- `hsGPA`: high school GPA
  - both college and high school GPAs are on a four-point scale.
- `ACT`: standardized test score
- `PC`: dummy variable
  - 1 if a student owns a personal computer
  - 0 otherwise

Q1

1. How many students own PC?

```
gpa1 %>%
  count(PC)
```

```
##   PC  n
## 1  0 85
## 2  1 56
```

Q2

Estimate the following model:

$$colGPA = \beta_0 + \beta_1\ hsGPA + \beta_2\ ACT + u.$$

```r
model_gpa <- lm(colGPA ~ hsGPA + ACT, data = gpa1)
model_gpa$coefficients
```

```
## (Intercept)       hsGPA         ACT
## 1.286327767 0.453455885 0.009426012
```

$$\widehat{colGPA} = 1.29 + 0.453\ hsGPA + \ 0.0094ACT$$

2. *There is a positive partial relationship between* `colGPA` *and* `hsGPA`*: holding ACT fixed, another point on* `hsGPA` *is associated with .453 of a point on the college GPA* — True or False?

True!

Q3

$$\widehat{colGPA} = 1.29 + 0.453 \ hsGPA + \ 0.0094ACT$$

3. *Given the fact that the highest possible score on the ACT is 36, the coefficient on ACT practically small.* — True or False?

True!

Q4

4. To determine the effects of computer ownership on college grade point average, estimate the following model:

$$colGPA = \beta_0 + \delta_0\ PC + \beta_1\ hsGPA + \beta_2\ ACT + u.$$

```
model_gpa2 <- lm(colGPA ~ PC + hsGPA + ACT, data = gpa1)
model_gpa2$coefficients
```

```
## (Intercept)          PC       hsGPA         ACT
## 1.263519813 0.157309205 0.447241653 0.008659005
```

$$\widehat{colGPA} = 1.26 + 0.157\ PC + 0.447\ hsGPA + 0.0087\ ACT + u.$$

This equation implies that a student who owns a PC has a predicted GPA about 1.6 points higher than a comparable student without a PC. — True or False?

False!

Q5

5. Generate a factor variable PC2: “Yes” if `PC = 0` otherwise “No”. Similarly, generate PC3 with `levels = c("Yes", "No")`.

```
gpa1 <- gpa1 %>%
  mutate(PC2 = factor(if_else(PC == 1, "Yes", "No"),
                             levels = c("Yes", "No")),
         PC3 = factor(if_else(PC == 1, "Yes", "No")))
```

Estimate the following models:

$$colGPA = \beta_0 + \delta_0 \ PC2 + \beta_1 \ hsGPA + \beta_2 \ ACT + u.$$
$$colGPA = \beta_0 + \delta_0 \ PC3 + \beta_1 \ hsGPA + \beta_2 \ ACT + u.$$

```r
model_gpa3 <- lm(colGPA ~ PC2 + hsGPA + ACT, data = gpa1)
model_gpa3$coefficients
```

```
##  (Intercept)        PC2No        hsGPA          ACT
##  1.420829018 -0.157309205  0.447241653  0.008659005
```

```r
model_gpa4 <- lm(colGPA ~ PC3 + hsGPA + ACT, data = gpa1)
model_gpa4$coefficients
```

```
## (Intercept)      PC3Yes       hsGPA         ACT
## 1.263519813 0.157309205 0.447241653 0.008659005
```

Questions?

bhaswar.chakma@ucp.pt