

Course 2 Section 3.6 - OPTIMISATION

Jiaying Wu

16/10/2020

```
#load library
library(tidyverse)
library(broom)
```

Give it a go

Continue to develop your understanding of the OLS estimator, optimal fitted line and residuals by making your way through this exercise. You'll need to use a pen or pencil and paper to produce your work, which you can then share with other learners.

Below is a data set with 5 observations for the two variables, x and y:

```
df <- tibble(x=c(1.5, 2, 3.1, 3.9, 5), y=c(3.7, 4.6, 5.5, 5.7, 7.2))
df
```

```
## # A tibble: 5 x 2
##       x     y
##   <dbl> <dbl>
## 1  1.5   3.7
## 2    2   4.6
## 3  3.1   5.5
## 4  3.9   5.7
## 5    5   7.2
```

Training a linear model of y using this data set produces the following equation:

```
fit_model <- lm(y ~ x, df)
tidy(fit_model)
```

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)    2.53      0.361     7.01 0.00595
## 2 x              0.905     0.108     8.39 0.00355
```

$$\hat{y} = 2.5338 + 0.9052x$$

Q1. compute the fitted values of y, i.e., the prediction of y for each value of x.

```
df <- df %>%
  mutate(y_hat = predict(fit_model))

df
```

```
## # A tibble: 5 x 3
##       x     y y_hat
##   <dbl> <dbl> <dbl>
## 1   1.5   3.7  3.89
## 2    2    4.6  4.34
## 3   3.1   5.5  5.34
## 4   3.9   5.7  6.06
## 5    5    7.2  7.06
```

Q2.use the predictions to compute the residuals and then the SSR

```
df <- df %>%
  mutate(residual = residuals(fit_model))

df
```

```
## # A tibble: 5 x 4
##       x     y y_hat residual
##   <dbl> <dbl> <dbl>   <dbl>
## 1   1.5   3.7  3.89  -0.192
## 2    2    4.6  4.34   0.256
## 3   3.1   5.5  5.34   0.16
## 4   3.9   5.7  6.06  -0.364
## 5    5    7.2  7.06   0.140
```

```
# SSR
sum(df$residual ^ 2)
```

```
## [1] 0.27998
```

Repeat the exercise above using the following less optimal fitted model:

$$\hat{y} = 2.5 + 1.1x$$

Q3.compute the fitted values of y

```
df2 <- tibble(x=c(1.5, 2, 3.1, 3.9, 5), y=c(3.7, 4.6, 5.5, 5.7, 7.2)) %>%
  mutate(y_hat = 2.5+1.1*x)

df2
```

```
## # A tibble: 5 x 3
##       x     y y_hat
##   <dbl> <dbl> <dbl>
## 1   1.5   3.7  4.15
```

```
## 2    2    4.6 4.7
## 3    3.1  5.5 5.91
## 4    3.9  5.7 6.79
## 5    5    7.2 8
```

Q4.use the predictions to compute the residuals and then the SSR

```
df2 <- df2 %>%
  mutate(residual = y - y_hat)

df2
```

```
## # A tibble: 5 x 4
##       x     y y_hat residual
##   <dbl> <dbl> <dbl>   <dbl>
## 1  1.5   3.7  4.15  -0.45
## 2    2    4.6  4.7   -0.1
## 3  3.1   5.5  5.91  -0.41
## 4  3.9   5.7  6.79  -1.09
## 5    5    7.2  8    -0.800
```

```
# SSR
sum(df2$residual ^ 2)
```

```
## [1] 2.2087
```