

STAT5002 Weekly Independent Exercises

Sheet 2 - Week 5

STAT5002

1 Linear model

The dataset `faithful` contains information about the waiting time between eruptions and the duration of each of the eruptions for the Old Faithful Geyser in Yellowstone National Park.

Using the following R outputs, write down a linear regression model to predict the value for `eruption` given a value of `waiting`. Round the intercept and the slope to two decimal places.

```
eruption = faithful$eruptions
waiting = faithful$waiting
c(round(mean(waiting), 2), round(sd(waiting), 2))
```

```
[1] 70.90 13.59
```

```
c(round(mean(eruption), 2), round(sd(eruption), 2))
```

```
[1] 3.49 1.14
```

```
round(cor(eruption, waiting), 2)
```

```
[1] 0.9
```

2 Coefficient of determination

Suppose we have a bivariate sample $(x_i, y_i), i = 1, \dots, n$ with $SD(x) = 1.5$ and $SD(y) = 0.5$. After fitting the linear regression model to the data, we obtain the regression line $y = a + bx$, where $a = 3$ and $b = 0.3$. What proportion of the total variation in the dependent variable y can be explained by the linear regression model?

3 Probability

Suppose a smoke-detector system consists of two parts A and B. If smoke occurs then A detects it with probability 0.95, B detects it with probability 0.98 and both of them detect it with probability 0.94. If smoke occurs, using the information given, solve the following tasks:

- (A) Write down $P(\text{A detects smoke})$, $P(\text{B detects smoke})$ and $P(\text{Both A and B detects smoke})$.
- (B) Show that the event “A detects smoke” and the event “B detects smoke” are not independent.
- (C) What is the probability that the smoke will not be detected by any of the sensors?
- (D) What is the probability that A will not detect the smoke, given that B did detect the smoke. – This is a challenging task (typically more difficult than exam questions you expect to see). You may need to apply the multiplication rule and the conditional probability.