STAT5002 Weekly Independent Exercises - solution

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
```

`Answer:`

The slope is

$$= r \times \frac{SD(\text{eruption})}{SD(\text{waiting})} = 0.9 \times \frac{1.14}{13.59} \approx 0.075$$

and the intercept is

$$a = mean(\text{eruption}) - b \times mean(waiting) \approx -1.86$$

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?

`Answer:`

To determine the proportion of total variation in y explained by the linear regression model, we need to compute the coefficient of determination, r^2.

To find the correlation coefficient r, we use the formula:

$$b = r \times \frac{SD(y)}{SD(x)}$$

which gives

$$r = b \times \frac{SD(x)}{SD(y)} = 0.3\frac{1.5}{0.5} = 0.9.$$

This way, $r^2 = 0.81$, so that 81% 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.

`Answer:`

(A) $P(\text{A detects smoke}) = 0.95$, $P(\text{B detects smoke}) = 0.98$ and $P(\text{Both A and B detects smoke}) = 0.94$.

(B) Two events are independent if:

$$P(\text{Both A and B detect smoke}) = P(\text{A detects smoke}) \times P(\text{B detects smoke})$$

However, we have

$$P(\text{A detects smoke}) \times P(\text{B detects smoke}) \approx 0.931$$

which is not $P(\text{Both A and B detect smoke}) = 0.94$, so they are not independent.

(C) The smoke is not detected if neither A nor B detects it, which is mutually exclusive to the event that either A or B detects it. So we have

$$P(\text{Neither A nor B detects smoke}) = 1 - P(\text{Either A or B detects smoke})$$

where $P(\text{Either A or B detects smoke}) =$

$$P(\text{A detects smoke}) + P(\text{B detects smoke}) - P(\text{Both A and B detects smoke}) = 0.99$$

by the addition rule. Thus,

$$P(\text{Neither A nor B detects smoke}) = 1 - P(\text{Either A or B detects smoke}) = 0.01.$$

(D) By the multiplication rule, we have

$$P(\text{A doesn't detect the smoke and B detects smoke}) =$$

$$P(\text{A doesn't detect the smoke}|\text{B detects smoke}) \times P(\text{B detects smoke})$$

so we have the conditional probability

$$P(\text{A doesn't detect the smoke}|\text{B detects smoke}) =$$

$$\frac{P(\text{A doesn't detect the smoke and B detects smoke})}{P(\text{B detects smoke})}.$$

Now we need to find out $P(\text{A doesn't detect the smoke and B detects smoke})$. Since we have

$$P(\text{A doesn't detect the smoke and B detects smoke}) +$$

$$P(\text{A detects the smoke and B detects smoke}) =$$

$$P(\text{B detects smoke regardless what A does}) = P(\text{B detects smoke})$$

which gives

$$P(\text{A doesn't detect the smoke and B detects smoke}) + \underbrace{P(\text{Both A and B detect smoke})}_{0.94} = 0.98$$

and thus $P(\text{A doesn't detect the smoke and B detects smoke}) = 0.04$. This gives

$$P(\text{A doesn't detect the smoke}|\text{B detects smoke}) = \frac{0.04}{0.98} \approx 0.0408.$$