

# 4.2 Correlation & Regression

## Question Paper

|            |                              |
|------------|------------------------------|
| Course     | DPIB Maths                   |
| Section    | 4. Statistics & Probability  |
| Topic      | 4.2 Correlation & Regression |
| Difficulty | Very Hard                    |

**Time allowed:** 90  
**Score:** /68  
**Percentage:** /100

**Question 1a**

What to Watch (WTW) and Bingeable are two organisations that review and rank television series. Based on different sets of criteria, scores out of 5 are assigned to 6 recent television series (labelled A to F). The data is shown in the table below.

| TV series                 | A   | B   | C   | D   | E   | F   |
|---------------------------|-----|-----|-----|-----|-----|-----|
| WTW'S score ( $x$ )       | 4.6 | 4.5 | 3.9 | 4.8 | 1.2 | 1.5 |
| Bingeable's score ( $y$ ) | 4.9 | 2.5 | 1.5 | 3.2 | 1.1 | 1.4 |
| WTW's rank                |     |     |     |     |     |     |
| Bingeable's rank          |     |     |     |     |     |     |

- (a) (i) Find the Pearson's product-moment correlation coefficient,  $r$ , for this data.
- (ii) Write down the equation of the regression line  $y$  on  $x$ .

[4 marks]

**Question 1b**

(b) Fill in the two empty rank rows in the table above.

[2 marks]

**Question 1c**

(c) Find the value of the Spearman's rank correlation coefficient,  $r_s$ .

[2 marks]

**Question 1d**

(d) Comment on the difference in the values of  $r$  and  $r_s$ .

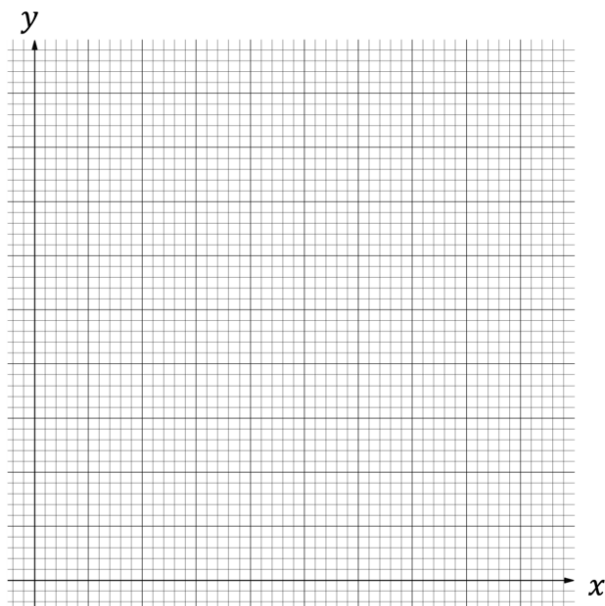
[2 marks]

**Question 2a**

The table below shows the lengths, in km, of 5 taxi rides in Melbourne, Australia and the corresponding prices, in AUD.

|                       |       |      |      |      |      |
|-----------------------|-------|------|------|------|------|
| Length, in km ( $x$ ) | 12.1  | 4.2  | 9.1  | 3.7  | 6.2  |
| Price, in AUD ( $y$ ) | 26.75 | 5.75 | 8.50 | 5.50 | 6.95 |

(a) Draw a scatter diagram for the above data on the axes below.



[4 marks]

**Question 2b**

(b) Calculate

- (i)  $\bar{x}$ , the mean taxi ride length
- (ii)  $\bar{y}$ , the mean price
- (iii) Plot the point  $M(\bar{x}, \bar{y})$  on your scatter diagram.

[3 marks]

**Question 2c**

- (c) (i) Write down the equation of the regression line  $y$  on  $x$ .
- (ii) Draw the regression line  $y$  on  $x$  on your scatter diagram.

[3 marks]

**Question 2d**

- (d) Show that the point  $M(\bar{x}, \bar{y})$  lies on the regression line  $y$  on  $x$ .

[1 mark]

**Question 3a**

A health study was conducted on 5 male and 5 female participants, measuring their average daily caffeine intake, in milligrams (mg), and their resting heart rate, in beats per minute (BPM). The following table shows the results of the study.

|   |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|
| Average daily caffeine intake, in mg – male ( $x_m$ )   | 222 | 312 | 211 | 190 | 120 |
| Resting heart rate, in BPM – male ( $y_m$ )             | 57  | 72  | 60  | 48  | 50  |
| Average daily caffeine intake, in mg – female ( $x_f$ ) | 202 | 411 | 254 | 81  | 52  |
| Resting heart rate, in BPM – female ( $y_f$ )           | 57  | 81  | 71  | 45  | 49  |

(a) Calculate the Pearson's product-moment correlation coefficient for,

- (i) the male participants,  $r_m$ ,
- (ii) the female participants,  $r_f$ .

[4 marks]

**Question 3b**

(b) Write down the equation of the regression line

(i)  $y_m$  on  $x_m$

(ii)  $y_f$  on  $x_f$ .

[4 marks]

**Question 3c**

(c) Find the intersection of the two regression lines found in part (b) and interpret its meaning.

[3 marks]

**Question 4a**

The following table shows the distance, in km, to 5 different ferry destinations from Rostock, Germany and the corresponding price of the cruise, in €.

| Destination             | Copenhagen | Oslo  | Stockholm | Helsinki | Riga   |
|-------------------------|------------|-------|-----------|----------|--------|
| Distance, in km ( $D$ ) | 174        | 620   | 730       | 933      | 810    |
| Price, in € ( $P$ )     | 30.50      | 65.00 | 45.75     | 85.50    | 125.00 |

The regression line  $P$  on  $D$  can be written in the form  $P = a + bD$ .

(a) Calculate the values of  $a$  and  $b$  and interpret their meanings

[3 marks]

**Question 4b**

The distance to Aberdeen from Rostock is 1093 km.

(b) Estimate the cost of the ferry to Aberdeen.

[2 marks]

**Question 4c**

(c) Comment on the reliability of your estimate found in part (b).

[1 mark]



**Question 4d**

The ferry company decides to charge €135 to travel to Aberdeen.

(d) Calculate the percentage error between your estimate found in part (b) and the actual price.

[3 marks]

**Question 5a**

The following table shows the total revenue,  $R$ , in £, obtained weekly during the first 7 weeks of 2021 by Larry, an independent financial consultant, and the number of clients,  $x$ , served.

| Week                  | 1    | 2    | 3    | 4    | 5    | 6    | 7    |
|-----------------------|------|------|------|------|------|------|------|
| Revenue, in £ ( $R$ ) | 2452 | 5751 | 6429 | 1203 | 4587 | 9786 | 6911 |
| Clients, $x$          | 7    | 11   | 14   | 4    | 5    | 8    | 9    |

(a) Write down the equation of the regression line  $R$  on  $x$ .

[2 marks]

**Question 5b**

Larry's weekly operating costs are £2250 and the cost of serving each client is £35.

(b) Find an expression for the profit Larry makes when serving  $x$  clients in a week.

[3 marks]

**Question 5c**

(c) Estimate the least number of clients required to generate a minimum of £1000 profit.

[3 marks]

**Question 6a**

Sandy Café is located on a beach and is open all year. The manager wants to see whether the daily average temperature, in °C, is correlated with the average tip they receive, as a percentage of the customer's total bill. He records this data over 9 days and details it in the table below.

|   |      |      |      |      |      |     |      |      |      |
|---|------|------|------|------|------|-----|------|------|------|
| Daily average temperature, in °C ( $x$ )              | 22.4 | 27.8 | 15.4 | 12.2 | 8.8  | 2.1 | 33.4 | 14.7 | 19.4 |
| Average tip as a percentage of the total bill ( $y$ ) | 20.1 | 16.3 | 12.4 | 12.8 | 10.1 | 9.4 | 18.8 | 13.1 | 15.9 |

(a) (i) Find the Pearson's product-moment correlation coefficient,  $r$ , for this data.

(ii) Write down the equation of the regression line  $y$  on  $x$ .

[2 marks]

**Question 6b**

On the 10<sup>th</sup> day, the average temperature is 25 °C and a customer tips their waiter \$20.

(b) Use the regression line to estimate the customer's total bill. Give your answer to 2 decimal places.

[4 marks]

**Question 6c**

The customer's total bill was \$98.50.

- (c) (i) Calculate the percentage error between your estimate found in part (b) and the actual total bill.
- (ii) Calculate the tip as a percentage of the actual total bill. Give your answer to the nearest whole number.

[4 marks]

**Question 7a**

The table below shows the petrol prices, in New Zealand dollars (NZD) per litre, for 6 different petrol stations (labelled A to F) along with their distance **south** of Auckland's city centre.

| Petrol station                            | A    | B    | C    | D    | E    | F    |
|---|------|------|------|------|------|------|
| Distance south of Auckland, in km ( $x$ ) | 122  | 314  | 456  | 231  | 178  | 392  |
| Petrol price, in NZD per litre ( $y$ )    | 1.94 | 1.88 | 1.78 | 1.84 | 1.99 | 1.81 |

(a) Calculate the mean petrol price,  $\bar{y}$ .

[1 mark]

**Question 7b**

The equation of the regression line  $y$  on  $x$  can be written in the form  $y = a + bx$ .

(b) (i) Calculate the value of  $a$ .

(ii) Calculate the value of  $b$ , giving your answer in the form  $k \times 10^n$ , where  $1 \leq |k| < 10, n \in \mathbb{Z}$ .

[3 marks]

**Question 7c**

The distance between Auckland's city centre and a new petrol station, G, is 200 km and the bearing of G from Auckland's city centre is  $166^\circ$ .

(c) Estimate the petrol price at G.

[2 marks]

**Question 7d**

Petrol station G decides to set their price at 0.44 NZD per litre below the mean petrol price calculated in part (a).

(d) Calculate the percentage error between the estimated petrol price for G calculated in part (c) and the actual petrol price at G.

[3 marks]