

# 4.2 Correlation & Regression

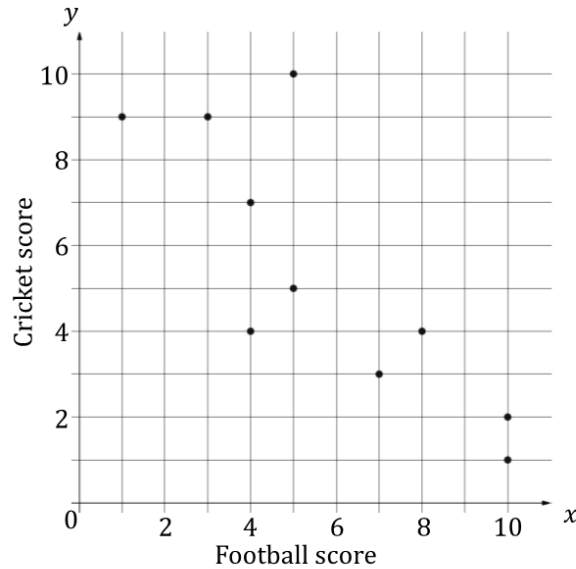
## Question Paper

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

**Time allowed:** 120  
**Score:** /96  
**Percentage:** /100

**Question 1a**

10 students are asked to give a score from 0 to 10 on how much they enjoy watching football and how much they enjoy watching cricket. The scores are shown in the scatter plot below.



(a) Use the scatter diagram to complete the missing cells in the table below.

Football score, $x$	1	3	4	4	5	5	7	8	10	10
Cricket score, $y$			4			10				2

[2 marks]

**Question 1b**

Another student only gave a cricket score of 6 and no football score.

(b) Estimate the football score for the student who has a cricket score of 6.

[2 marks]

**Question 1c**

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

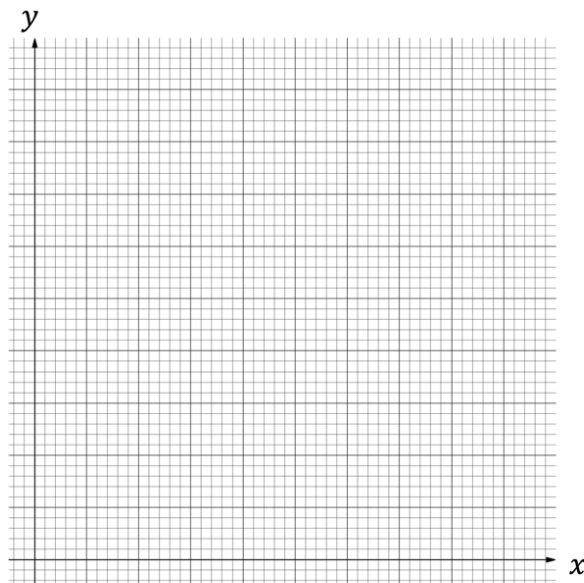
[1 mark]

**Question 2a**

A study is conducted on 6 participants (labelled A to F) measuring their average number of hours of sleep per night and their score, from 0 to 100, in a short-term memory test. The results of the study are shown in the table below.

Participant	A	B	C	D	E	F
Average number of hours of sleep ( $x$ )	6.8	7.2	8.1	9.4	5.9	7.5
Short term memory test score ( $y$ )	72	70	82	79	62	80

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



[4 marks]

**Question 2b**

- (b) (i) Calculate the Pearson's product-moment correlation coefficient,  $r$ .
- (ii) Write down the equation of the regression line  $y$  on  $x$ .
- (iii) Draw the regression line on your scatter diagram.

**[3 marks]****Question 2c**

- (c) Use the regression line from part (b) to estimate the average number of hours of sleep a participant gets per night when their score in the memory test is 67. Give your answer to the nearest integer.

**[2 marks]**

**Question 2d**

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

[1 mark]

**Question 3a**

The following table shows the total CO<sub>2</sub> emissions,  $T$  tonnes, from 5 different countries (labelled A to E) and their average annual household income,  $S$  USD.

Country	A	B	C	D	E
CO <sub>2</sub> emissions, $T$ tonnes	10 500 000	5 500 000	2 500 000	1 600 000	1 200 000
Average annual household income, $S$ USD	55 000	105 000	15 000	55 000	40 000

(a) (i) Calculate the Pearson's product-moment correlation coefficient,  $r$ .

(ii) Hence comment on the result.

[3 marks]

**Question 3b**

(b) Write down the equation of the regression line  $S$  on  $T$ .

[2 marks]

### Question 3c

(c) State two reasons why it would be inappropriate to use the regression line from part (b) to estimate the percentage of total global emissions from a country where the average household annual income is 75 000 USD.

[2 marks]

### Question 4a

Sweet Dreams is a company that hotel ratings website that personally rates hotels on a scale of 1 to 10 along with allowing the public to leave ratings on the hotels. The ratings from the public include a comment section. The table below shows the ratings from Sweet dreams and the public for 9 different hotels (labelled A to I).

Hotel	A	B	C	D	E	F	G	H	I
Sweet Dream's rating	7	9	4	5	8	1	3	6	2
Public's rating	7.8	9.1	2.1	2.4	9.8	1.4	1.9	5.5	1.5
Sweet Dream's rank									
Public's rank									

(a) Complete the two empty rank rows in the table above, ranking the highest scores first.

[2 marks]

**Question 4b**

- (b) (i) Calculate the value of the Pearson's product-moment correlation coefficient,  $r$ .
- (ii) Calculate the value of the Spearman's rank correlation coefficient,  $r_s$ .

**[3 marks]****Question 4c**

A mistake was made when calculating the rating from the public for Hotel D. The real rating for Hotel D from the public is 2.3.

- (c) Explain why  $r_s$  does not change when the public's rating for Hotel D is changed.

**[1 mark]**

**Question 5a**

Easy Breezes is a company based in the snowy mountains of Greenland that makes heat pumps. Easy Breezes wants to see if the average weekly temperature, in  $^{\circ}\text{C}$ , is correlated with the average weekly energy consumption from their air conditioning units, in kilowatt hours (kWh). Easy Breezes records the following data.

Average weekly temperature, in $^{\circ}\text{C}$ ( $x$ )	-8.2	-4.3	-1.7	0.8	2.0	4.4	8.5
Average weekly energy consumption, in kWh ( $y$ )	365.2	316.4	292.7	249.1	187.2	142.8	131.2

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

**[4 marks]****Question 5b**

- (b) (i) Use your regression line from part (a) (i) to estimate the number of kWh one of Easy Breezes air conditioning units would use in a week when the weekly average temperature is  $11^{\circ}\text{C}$ .
- (ii) Comment on the reliability of your estimate.

**[4 marks]**



**Question 5c**

The actual usage of one of their heat pumps in a week where the average temperature was  $11^{\circ}\text{C}$  is 52.0 kWh.

(c) Calculate the percentage error in your estimate found in part (b) (i) and the actual usage.

[2 marks]

**Question 6a**

A supermarket has a physical and online store. The following table shows the total daily revenue,  $y$ , in USD, along with the number of customers that they had come into their physical store during the day, over 8 separate days.

Customers ( $x$ )	45	88	54	97	154	101	36	72
Revenue, USD ( $y$ )	548.21	832.55	497.71	1021.97	1138.73	988.62	1026.21	754.38

- (a) (i) Calculate the Pearson's product-moment correlation coefficient,  $r$ .
- (ii) Hence comment on the result.

[3 marks]

**Question 6b**

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

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

[3 marks]

**Question 6c**

The supermarket has daily fixed costs of 650 USD.

(c) Find an expression for the daily profit of the supermarket when they have  $x$  customers on a particular day.

[2 marks]

**Question 6d**

(d) Estimate the least number of physical customers required in order to make a profit on any particular day.

[2 marks]

**Question 7a**

10 rugby players (labelled A to J) are used to investigate the relationship between a player's maximum sprint velocity, in  $\text{ms}^{-1}$ , and their weight, in kg. The data is recorded in the table below.

Player	A	B	C	D	E	F	G	H	I	J
Weight, in kg ( $x$ )	96	99	88	95	112	98	85	108	82	109
Maximum sprint velocity, in $\text{ms}^{-1}$ ( $y$ )	7.5	6.9	10.1	8.8	6.1	6.9	5.8	10.7	11.9	6.5

(a) Calculate the value of the Pearson's product-moment correlation coefficient,  $r$

- (i) with players G and H.
- (ii) without the players G and H.

[4 marks]

**Question 7b**

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

- (i) with players G and H.
- (ii) without the players G and H.

[4 marks]

**Question 7c**

(c) Comment on the results found in part (a) and (b) and state whether you would use the regression line with players G and H or the regression line without players G and H when estimating a rugby player's maximum sprint velocity, given their weight.

[2 marks]

**Question 8a**

A study was conducted on 6 students measuring their arm length,  $x$  cm, and the maximum number of push ups they can do in a minute. The results of the study are shown in the table below.

Arm length, $x$ cm	72.2	69.2	75.6	78.1	78.5	74.5
No. of push ups, $y$	34	42	24	30	38	31

(a) State the range of the number of push ups.

[1 mark]

**Question 8b**

- (b) (i) Calculate the Pearson's product-moment correlation coefficient,  $r$ .
- (ii) Comment on the correlation between the athlete's arm length and the maximum number of push ups they can do in a minute.

[3 marks]

**Question 8c**

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

[2 marks]

**Question 8d**

Another student, Tom, is a sportsman and can do 62 push ups in one minute.

(d) Use the regression line found in part (c) to estimate Tom's arm length.

[2 marks]

**Question 8e**

(e) State whether your estimate is valid and justify your answer.

[2 marks]

**Question 9a**

Body mass index (BMI) is used to measure whether someone is over or under weight, however BMI does not take someone's body fat percentage into account. The following table shows the BMI and body fat percentage from 7 male participants.

BMI, $x$	22.4	19.8	25.5	29.8	31.2	18.1	17.2
Body fat %, $y$	22.1	20.1	24.2	31.1	16.2	15.1	8.6

(a) (i) Calculate the Pearson's product-moment correlation coefficient,  $r$ .

(ii) Comment on the result found for  $r$ .

[3 marks]

**Question 9b**

The regression line  $y$  on  $x$  is in the form  $y = mx + c$ .

(b) Calculate the values of  $m$  and  $c$  and interpret their meanings. Explain whether they are appropriate within the context of the question.

[4 marks]

**Question 9c**

The formula to calculate someone's BMI is

$$\text{BMI} = \frac{\text{weight in kilograms}}{(\text{height in metres})^2}$$

John weighs 95 kilograms and is 1.84 metres tall.

(c) Estimate John's body fat percentage and comment on the reliability of your estimate.

[3 marks]

**Question 10a**

A movie cinema is considering significantly reducing the price of their popcorn as they believe their customers spend more on drinks when they buy popcorn. They recorded the following data of the daily revenue from popcorn,  $\$x$ , and the daily revenue from drinks,  $\$y$  over 8 randomly selected days.

Popcorn revenue, $\$x$	78.20	102.50	30.80	22.20	132.90	200.50	154.80	132.40
Drinks revenue, $\$y$	202.10	308.50	60.70	75.80	270.50	300.00	368.20	198.70

(a) Calculate

- (i)  $\bar{x}$ , the mean daily revenue from popcorn.
- (ii)  $\bar{y}$ , the mean daily revenue from drinks.
- (iii)  $r$ , the Pearson's product-moment correlation coefficient.

[4 marks]



**Question 10b**

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

- (b) Calculate the values of  $a$  and  $b$  and interpret their meanings and explain whether they are appropriate within the context of the question.

[4 marks]

**Question 10c**

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

[2 marks]

**Question 11a**

Lara, a biology teacher, wants to find out how much of the course her students have retained. She gives her students a practice exam and records their scores out of 100. In order to test the reliability of her exam she gives her students the same exam the following day. The results are shown below.

Student	A	B	C	D	E	F	G	H	I	J	K
1 <sup>st</sup> Score	71	65	43	51	87	72	50	93	42	31	84
2 <sup>nd</sup> Score	73	89	79	72	88	72	61	99	82	92	85

(a) State the name of this test for reliability.

[1 mark]

**Question 11b**

(b) Find Pearson's product moment correlation coefficient for the data in the table above.

[2 marks]

**Question 11c**

(c) Explain why Pearson's product moment correlation coefficient might not be a good indication of the reliability of Lara's exam.

[2 marks]

**Question 11d**

Lara thinks that the scores from the first exam can be used to predict the results in the end of term exam which has different questions. The results are shown below.

Student	A	B	C	D	E	F	G	H	I	J	K
End of term score	73	68	42	56	88	70	55	95	39	39	79

(d) Use a linear regression model to predict the end of term score for a student who gets a score of 60 on Lara's exam.

[3 marks]