

QQQ – Statistics Year 1 – Chapter 4 – Correlation (Version 1)

Total Marks: 23

(23 = Platinum, 21 = Gold, 18 = Silver, 16 = Bronze)

1. Manny is studying the price and number of pages of a random sample of books.

He calculates the value of the product moment correlation coefficient between the price and number of pages in each book as 1.05

Which of the following best describes the value 1.05?

Tick (✓) **one** box.

[1 mark]

- definitely correct
- probably correct
- probably incorrect
- definitely incorrect

(Total 1 marks)

AQA June 2019 Paper 2 Q12

2. A company is introducing a job evaluation scheme. Points (x) will be awarded to each job based on the qualifications and skills needed and the level of responsibility. Pay (£ y) will then be allocated to each job according to the number of points awarded.

Before the scheme is introduced, a random sample of 8 employees was taken and the linear regression equation of pay on points was $y = 4.5x - 47$

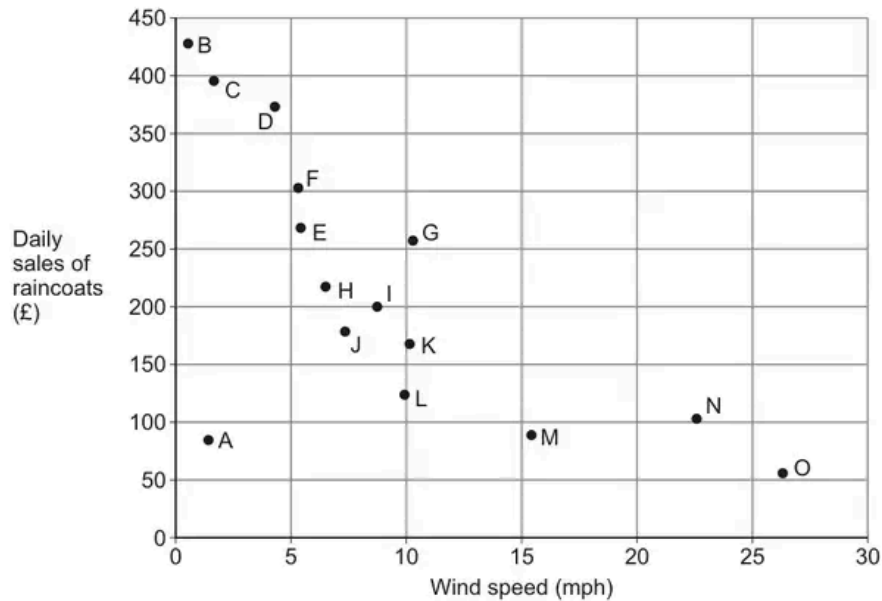
- (a) Describe the correlation between points and pay. **(1)**
- (b) Give an interpretation of the gradient of this regression line. **(1)**
- (c) Explain why this model might not be appropriate for all jobs in the company. **(1)**

(Total 3 marks)

Edexcel June 2018 Statistics Q1

3. The owner of a shop recorded the average daily wind speed in mph and the total sales of raincoats in pounds over a period of 6 months.

The data from a random selection of 15 days from the 6-month period is shown in the scatter diagram.



One of the points on the scatter diagram is an outlier.

- (a) State the letter from the diagram which corresponds to this outlier. **[1 mark]**
- (b) State the effect on the numerical value of the correlation coefficient if the outlier was removed from the data. **[1 mark]**
- (c) Describe, in context, the correlation indicated by the scatter diagram. **[1 mark]**
- (d) The owner of the shop claims that the scatter diagram proves that windier days cause lower sales of raincoats.

Comment on the validity of this claim.

[1 mark]

(Total 4 marks)

4. Fred and Nadine are investigating whether there is a linear relationship between Daily Mean Pressure, p hPa, and Daily Mean Air Temperature, t °C, in Beijing using the 2015 data from the large data set.

Fred randomly selects one month from the data set and draws the scatter diagram in Figure 1 using the data from that month.

The scale has been left off the horizontal axis.

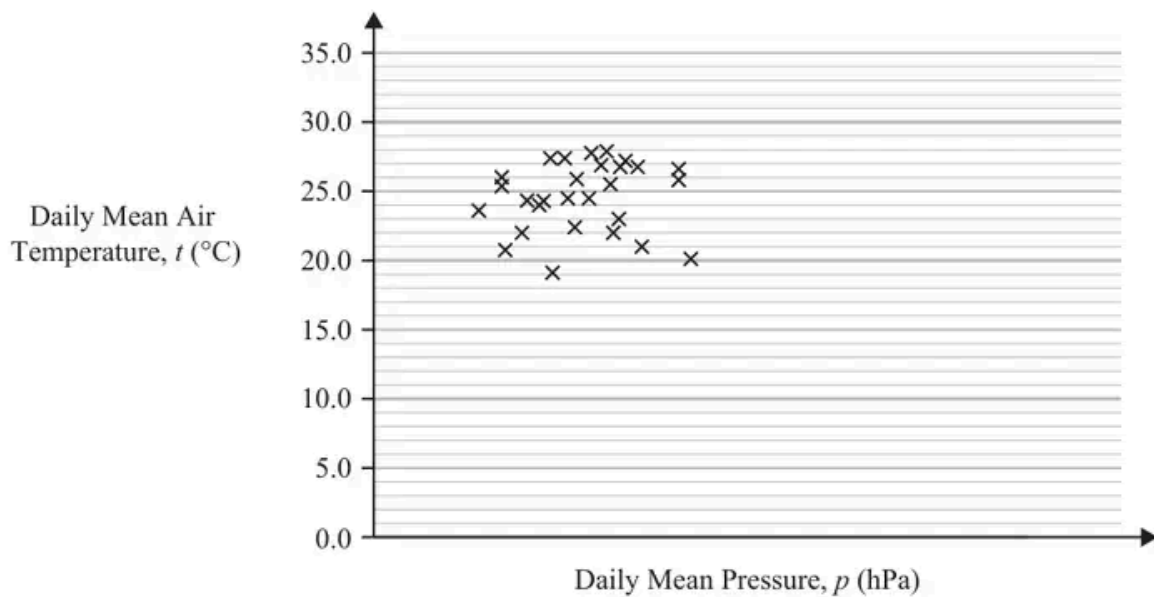


Figure 1

(a) Describe the correlation shown in Figure 1. (1)

Nadine chooses to use all of the data for Beijing from 2015 and draws the scatter diagram in Figure 2.

She uses the same scales as Fred.

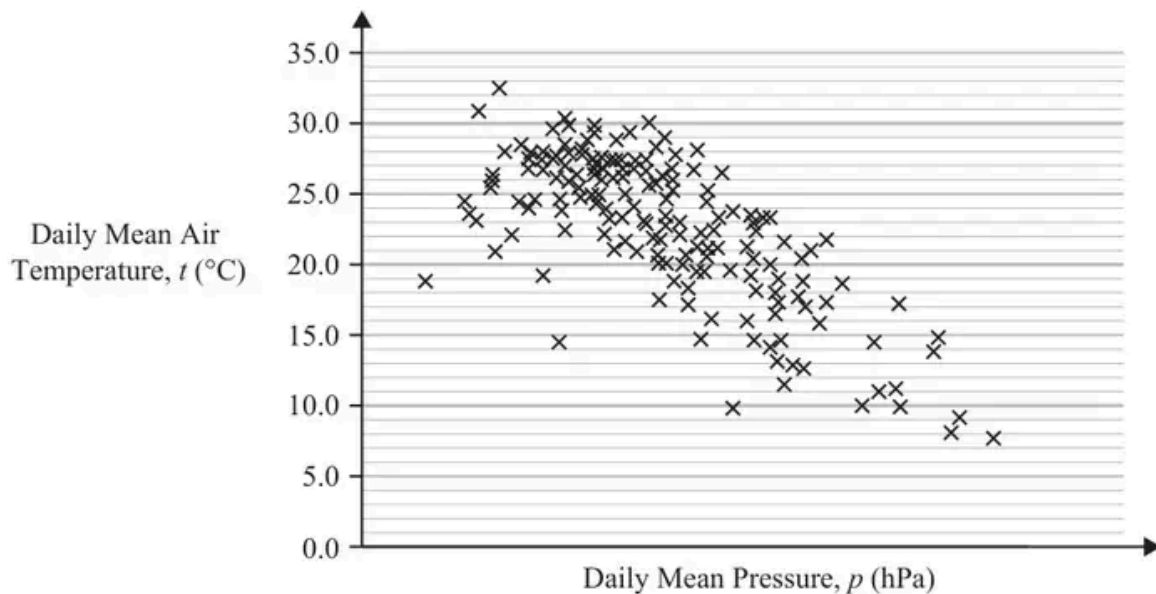


Figure 2

- (b) Explain, in context, what Nadine can infer about the relationship between p and t using the information shown in Figure 2. (1)
- (c) Using your knowledge of the large data set, state a value of p for which interpolation can be used with Figure 2 to predict a value of t . (1)
- (d) Using your knowledge of the large data set, explain why it is not meaningful to look for a linear relationship between Daily Mean Wind Speed (Beaufort Conversion) and Daily Mean Air Temperature in Beijing in 2015. (1)

(Total 4 marks)

5. The relationship between two variables p and t is modelled by the regression line with equation

$$p = 22 - 1.1 t$$

The model is based on observations of the independent variable, t , between 1 and 10

(a) Describe the correlation between p and t implied by this model. (1)

Given that p is measured in centimetres and t is measured in days,

(b) state the units of the gradient of the regression line. (1)

Using the model,

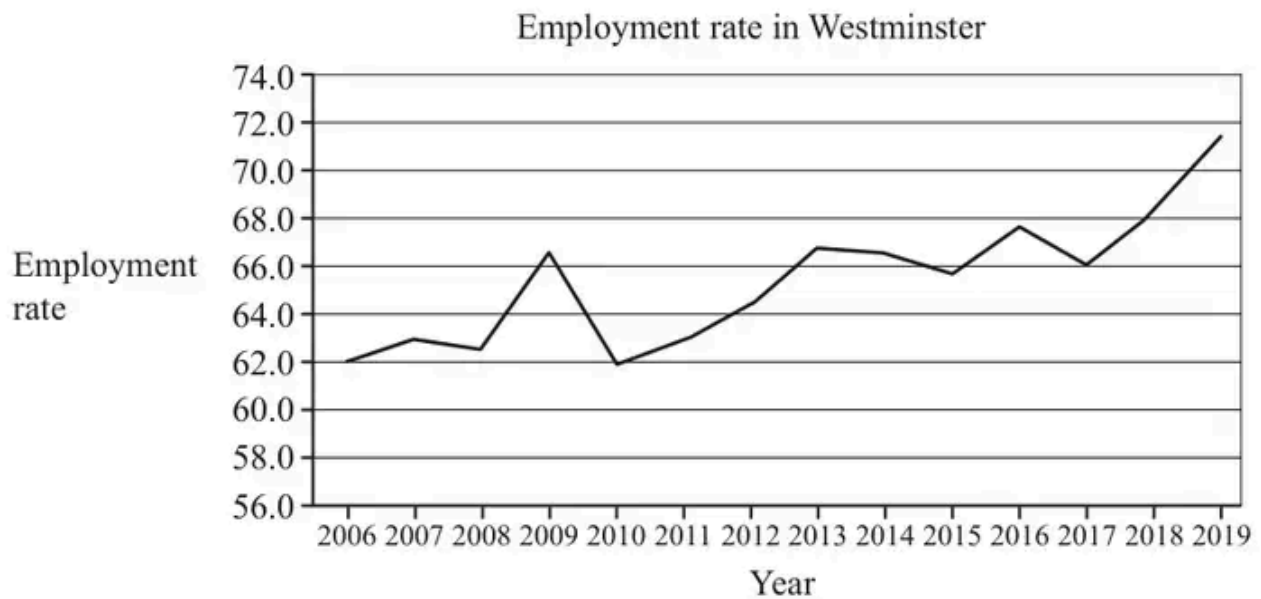
(c) calculate the change in p over a 3-day period. (2)

Tisam uses this model to estimate the value of p when $t = 19$

(d) Comment, giving a reason, on the reliability of this estimate. (1)

(Total 5 marks)

6. The pre-release material contains information about employment rates in London boroughs. The graph shows employment rates for Westminster between 2006 and 2019.



A local politician stated that the diagram shows that more than 60% of seventy-year-olds were in employment throughout the period from 2006 to 2019.

- (a) Use your knowledge of the pre-release material to explain whether there is any evidence to support this statement. [1]

In order to estimate the employment rate in 2020, two different models were proposed using the LINEST function in a spreadsheet.

Model 1 (using all the data from 2006 onwards)

$$Y = 0.549x - 1040,$$

Model 2 (using data from 2017 onwards)

$$Y = 2.65x - 5280,$$

where Y = employment rate and x = calendar year.

It was subsequently found that the employment rate in Westminster in 2020 was 68.4%.

- (b) Determine which of the two models provided the better estimate for the employment rate in Westminster in 2020. [3]
- (c) Use your knowledge of the pre-release material to explain whether it would be appropriate to use either model to estimate the employment rate in 2020 in other London boroughs. [1]
- (d) What does model 2 predict for employment rates in Westminster in the long term? [1]

(Total 6 marks)

OCR MEI June 2022 Paper 2 Q6