

QQQ – Statistics Year 1 – Chapter 4 – Correlation (Version 2)

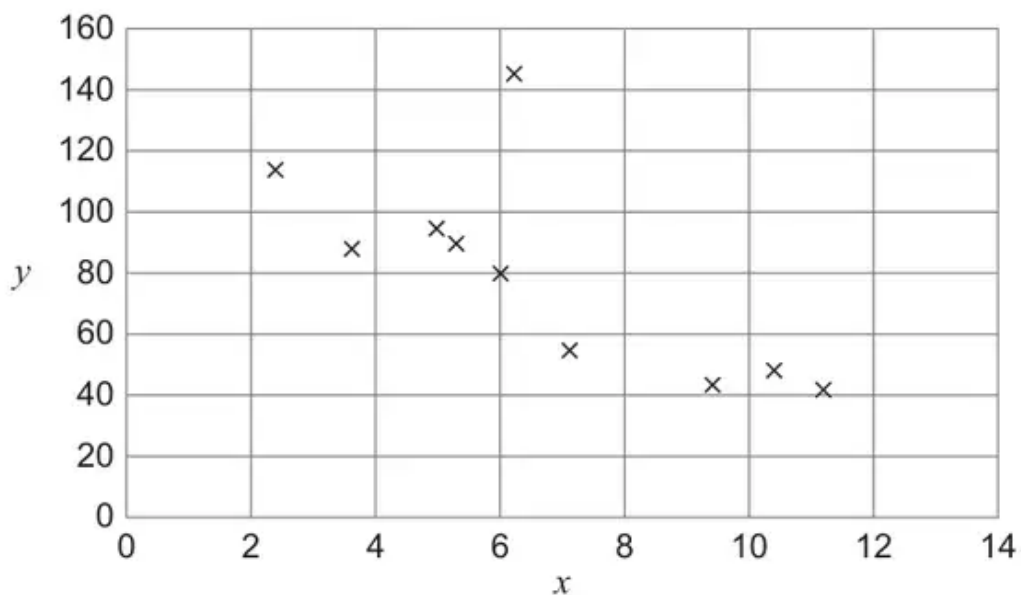
Total Marks: 22

(22 = Platinum, 20 = Gold, 18 = Silver, 15 = Bronze)

1. The table below shows the daily salt intake, x grams, and the daily Vitamin C intake, y milligrams, for a group of 10 adults.

Adult	A	B	C	D	E	F	G	H	I	J
x	5.3	6.2	3.6	10.4	2.4	9.4	6	5	7.1	11.2
y	90	145	88	48	114	44	80	95	55	41

A scatter diagram of the data is shown below.



One of the adults is an outlier. Identify the letter of the adult that is the outlier.

Circle your answer below.

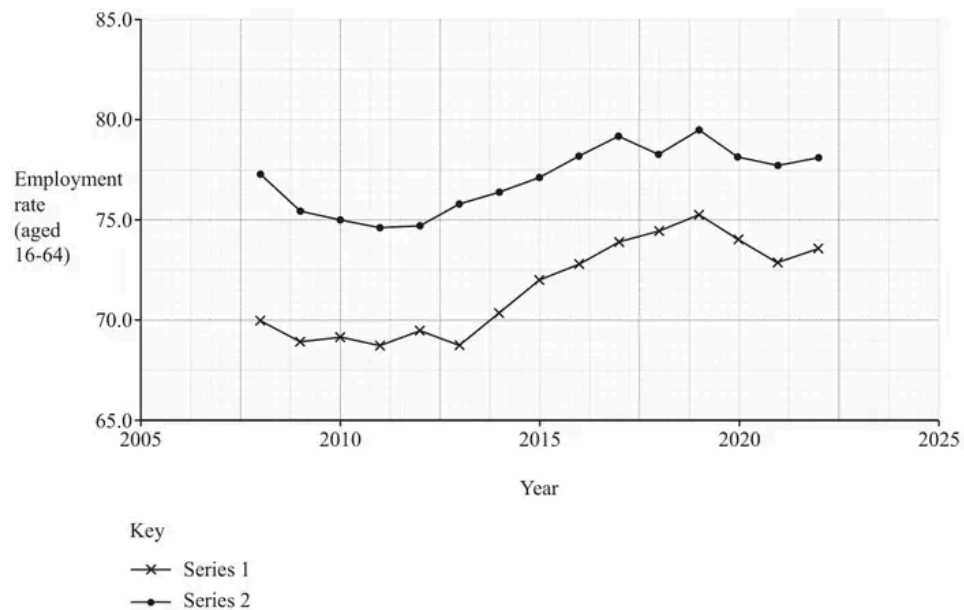
[1 mark]

- A
- B
- E
- J

(Total 1 marks)

2. The pre-release material contains data for percentage employment rates in different areas of England.

The diagram shows percentage employment rates between 2008 and 2022 for the North West of England and the South East of England.



- (a)** Use your knowledge of the pre-release material to explain whether Series 1 (the lower time series) or Series 2 (the upper time series) represents the data for the South East of England. [1]

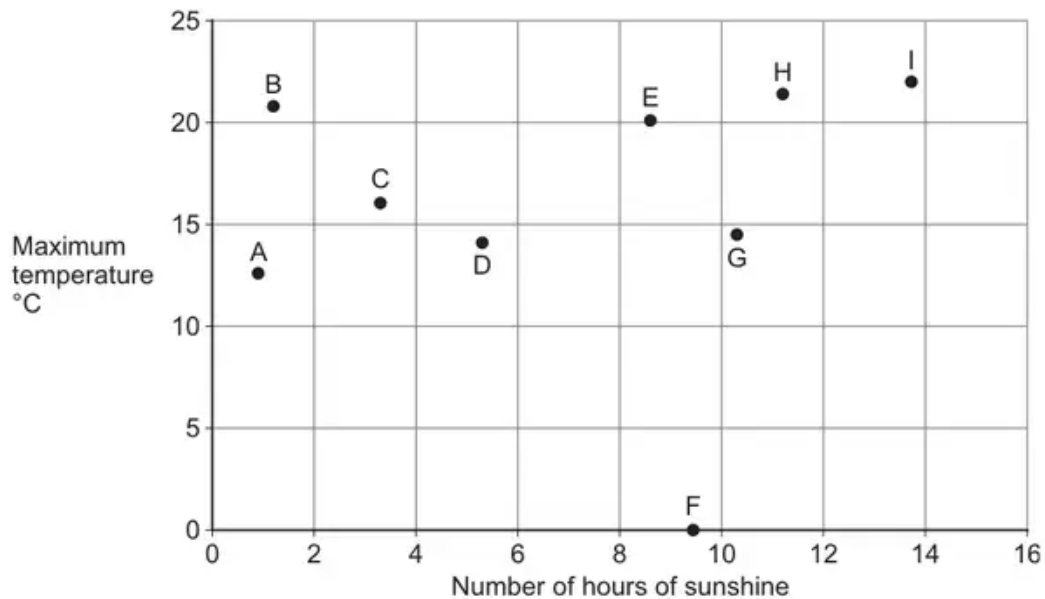
A student uses the LINEST function to model the data for Series 2.

- (b)** Assuming that a straight line model is reasonable, use your knowledge of the pre-release material to explain whether it would be appropriate to use this model to calculate an estimate of the percentage employment rate in Cornwall, in the South West of England, in 2011. [1]

(Total 2 marks)

3. The number of hours of sunshine and the daily maximum temperature were recorded over a 9-day period in June at an English seaside town.

A scatter diagram representing the recorded data is shown below.



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

- (a) (i) Write down the letter that identifies this point. [1 mark]
- (ii) Suggest one possible action that could be taken to deal with this error. [1 mark]
- (b) It is claimed that the scatter diagram proves that longer hours of sunshine cause higher maximum daily temperatures.

Comment on the validity of this claim.

[1 mark]

(Total 3 marks)

4. The pre-release material contains information for countries in the world concerning real GDP per capita in US\$ and mobile phone subscribers per 100 population. In an investigation into the relationship between these two variables, a student takes a sample of 20 countries in Africa. The student draws a scatter diagram for the data, which is shown in **Fig. 5.1**.

Africa 1st sample

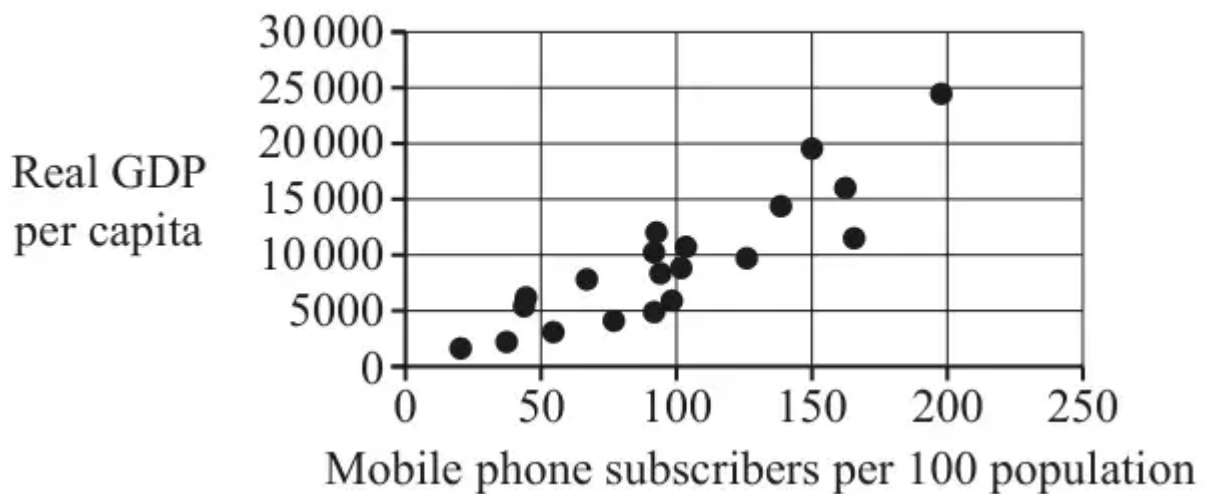


Fig. 5.1

- (a) What does **Fig. 5.1** suggest about the relationship between real GDP per capita and the number of mobile phone subscribers per 100 population? [1]

Another student collects a **different** sample of 20 countries from Africa, and draws a scatter diagram for the data, which is shown in **Fig. 5.2**.

Africa 2nd sample

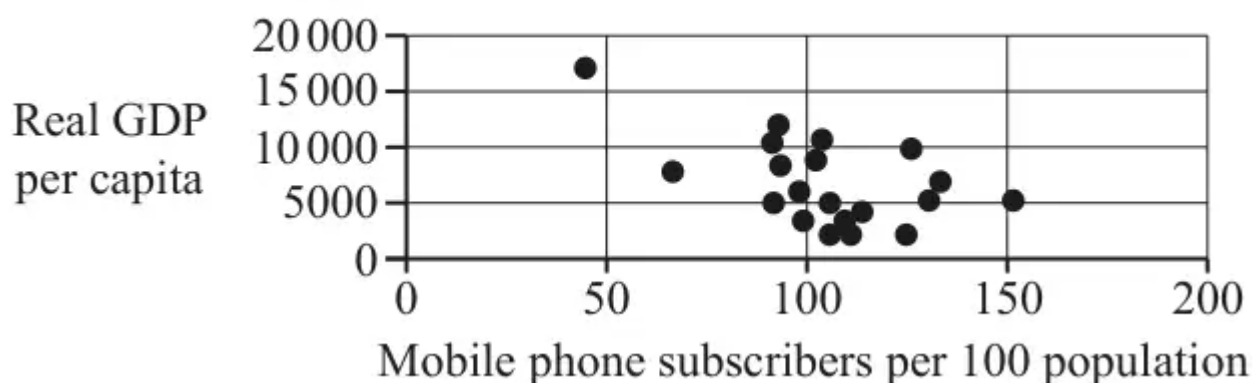


Fig. 5.2

- (b) What does **Fig. 5.2** suggest about the relationship between real GDP per capita and the number of mobile phone subscribers per 100 population? [1]
- (c) Explain whether either of the two scatter diagrams is likely to be representative of the true relationship between real GDP per capita and the number of mobile phone subscribers per

100 population, for countries in Africa.

[1]

(Total 3 marks)

OCR MEI June 2024 Paper 2 Q5

5. This question deals with information about the populations of Local Authorities (LAs) in the North of England, taken from the 2011 census.

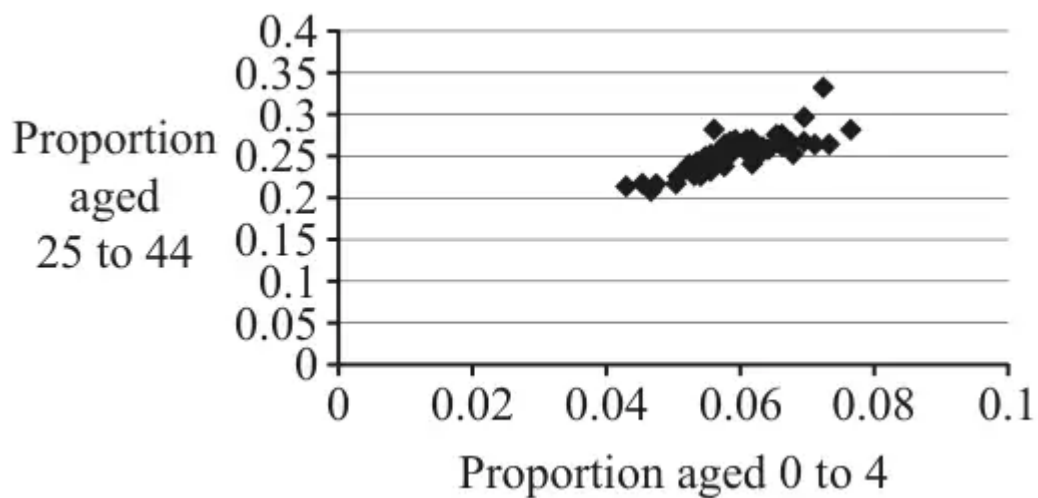


Fig. 1

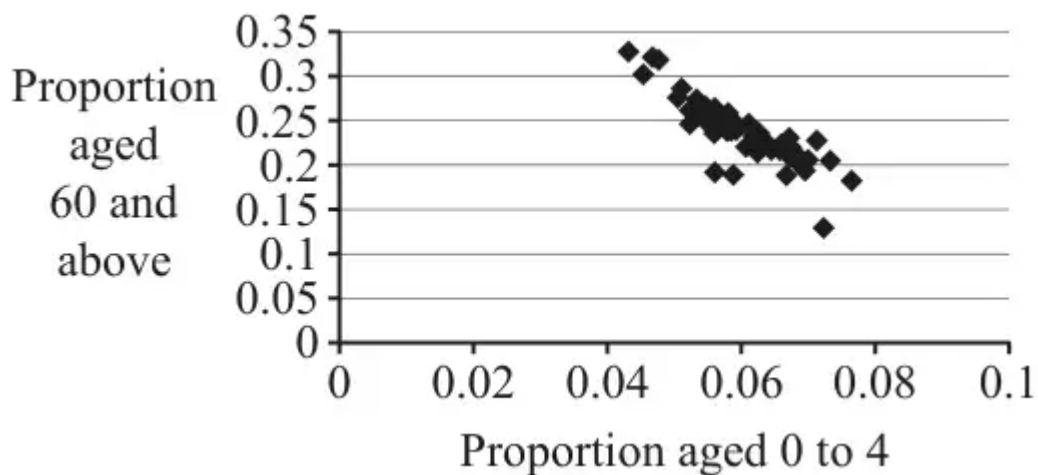


Fig. 2

Fig. 1 and **Fig. 2** both show strong correlation, but of two different kinds.

- (a) For each diagram, use a single word to describe the kind of correlation shown. [1]
- (b) For each diagram, suggest a reason, in context, why the correlation is of the particular kind described in part (a). [2]

Fig. 3 is the same as **Fig. 2** but with the point *A* marked.

Fig. 4 shows information about the same LAs as **Fig. 2** and **Fig. 3**.

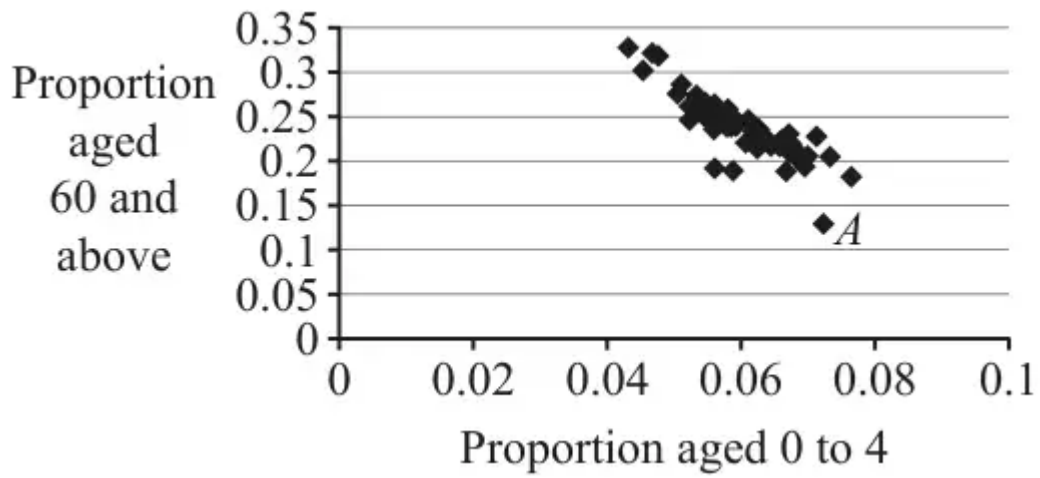


Fig. 3

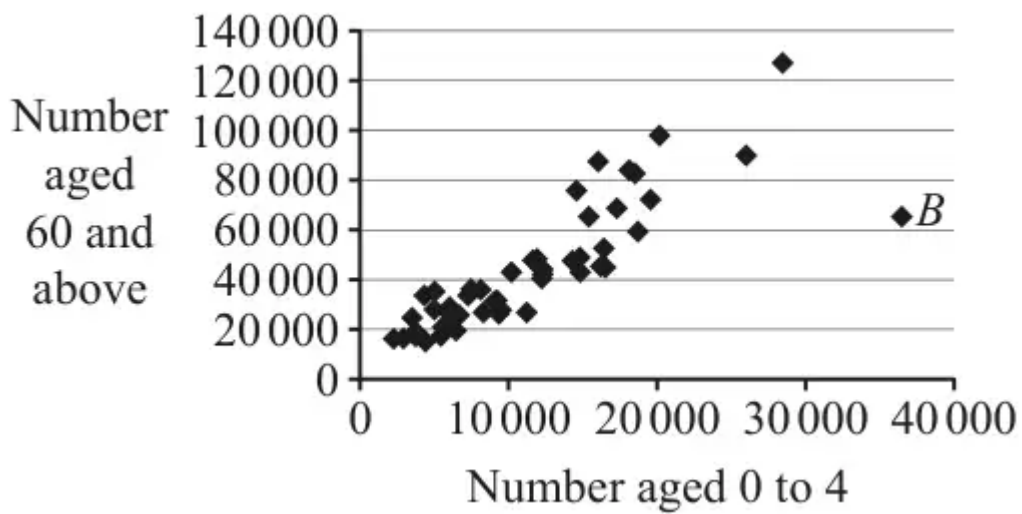


Fig. 4

- (c) Point *A* in Fig. 3 and point *B* in Fig. 4 represent the same LA.

Explain how you can tell that this LA has a large population.

[1]

(Total 4 marks)

6. The pre-release material contains information on Pulse Rate and Body Mass Index (BMI). A student is investigating whether there is a relationship between pulse rate and BMI. A **section** of the available data is shown in the table.

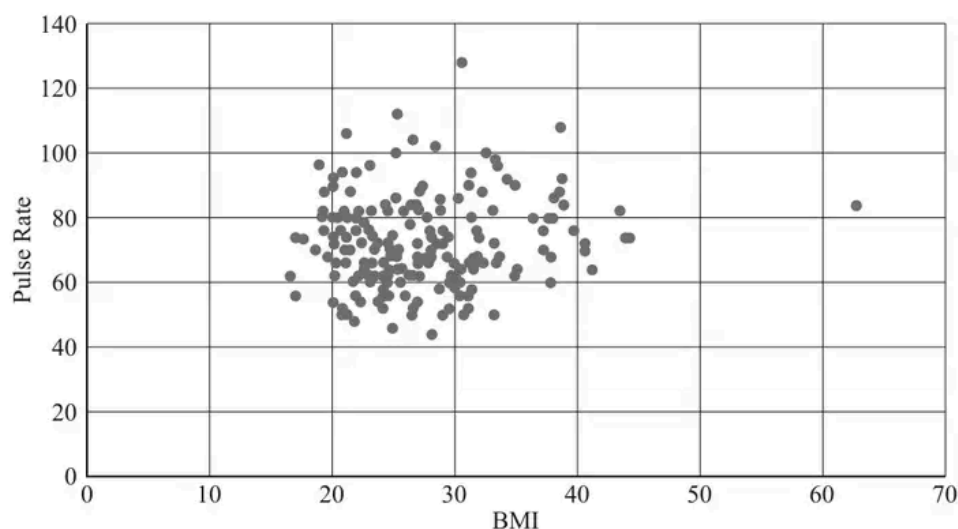
Sex	Age	BMI	Pulse
Male	62	29.54	60
Female	20	23.68	#N/A
Male	17	26.97	72
Male	35	24.7	64
Male	17	20.09	54
Male	85	23.86	54
Female	81	24.04	#N/A

The student decides to draw a scatter diagram.

- (a) With reference to the table, explain which data should be cleaned before any analysis takes place. [1]

The student cleans the data for BMI and Pulse Rate in the pre-release material and draws a scatter diagram.

Scatter diagram of Pulse Rate against BMI



The student identifies **one** outlier.

- (b) On the copy of the scatter diagram in the Printed Answer Booklet, circle this outlier. [1]

The student decides to remove this outlier from the data. They then use the LINEST function in the spreadsheet to obtain the following formula for the line of best fit.

$$P = 0.29Q + 64.2,$$

where P = Pulse Rate and Q = BMI.

They use this to estimate the Pulse Rate of a person with BMI 23.68.

They obtain a value of 71 correct to the nearest whole number.

- (c) With reference to the scatter diagram, explain whether it is appropriate to use the formula for the line of best fit. [1]

It is suggested that all pairs of values where the pulse rate is above 100 should also be cleaned from the data, as they must be incorrect.

- (d) Use your knowledge of the pre-release material to explain whether or not all pairs of values with a pulse rate of more than 100 should be cleaned from the data. [1]

(Total 4 marks)

7. Giovanni believes there is a relationship between the mass of a car, m kg, and its fuel consumption, c miles per gallon (mpg).

For a sample of 25 cars, he obtains the following summary statistic.

$$\sum (c - \bar{c})^2 = 394$$

(a) Find the standard deviation of the fuel consumption of the 25 cars. (1)

Using m as the explanatory variable, Giovanni creates the linear regression model

$$c = a + bm$$

where a and b are constants.

Using his model, he concludes that, on average

- the fuel consumption is 3.5 mpg lower for each additional 500 kg of mass
- the fuel consumption of a car with a mass of 1700 kg is 20 mpg

(b) (i) Find the value of b (2)

(ii) Find the value of a (2)

(Total 5 marks)