Statistics Year 1 (AS) exam questions

Arancha Ruiz

Ch. 4: Correlation

June 2023 Question 2

2. 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.

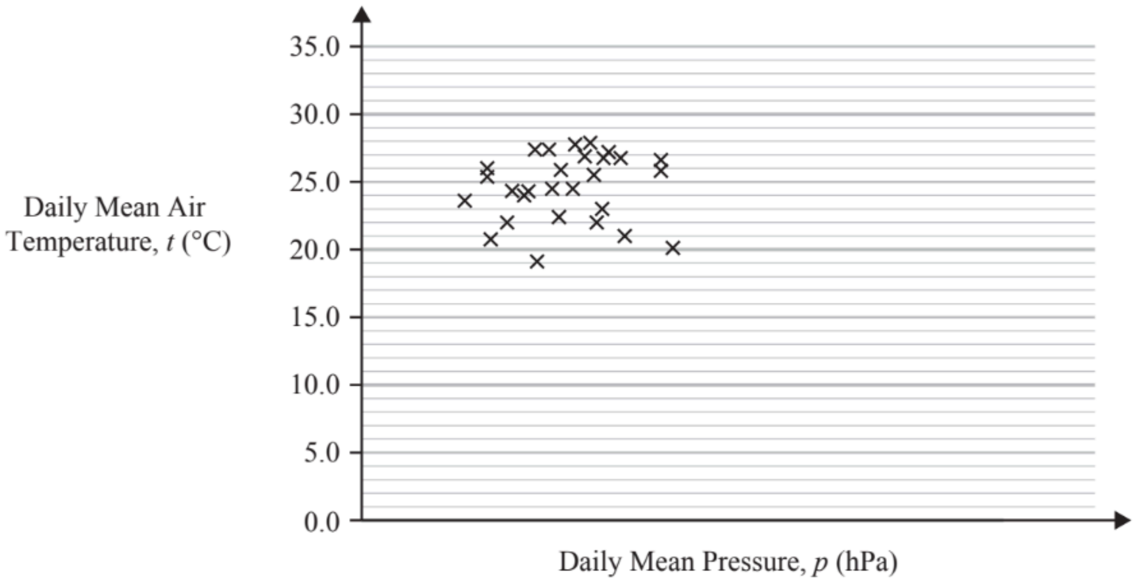

Figure 1

(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.

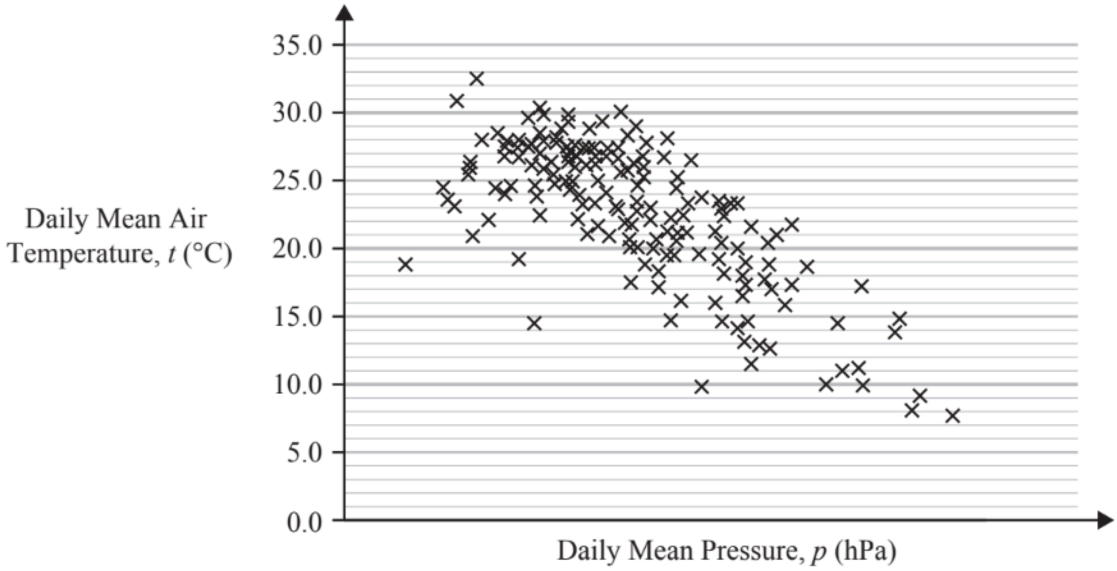

Figure 2

(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)

ANSWER

Question	Scheme	Marks	AOs
2(a)	No (correlation)/weak (correlation)	B1	1.1b
		(1)	
(b)	(Negative correlation…) As p(ressure) increases, t(emperature) decreases.	B1	2.2b
		(1)	
(c)	990 to 1040 (hPa)	B1	3.4 LDS
		(1)	
(d)	Daily mean wind speed (Beaufort) is a qualitative variable.	B1	2.4 LDS
		(1)	
			(4 marks)

June 2022 Question 1

1. 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)

ANSWER

Qu	Scheme	Marks	AO
1. (a)	Negative (since gradient of regression line is negative)	B1	1.2
		(1)	
(b)	cm/day (o.e. e.g. cm day^{-1})	B1	2.2a
		(1)	
(c)	$3\times[\pm]1.1$	M1	3.4
	= decrease of 3.3 [cm]	A1	1.1b
		(2)	
(d)	19 is (well) outside the range [1, 10] or involves extrapolation (o.e.) so (possibly) unreliable/ inaccurate (o.e.)	B1	2.4
		(1)	
		(5 marks)	

Video solution:

https://youtu.be/ZtQSYJYy51A

October 2020 question 2

2. Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

Date	**Daily Maximum Relative Humidity**	**Daily Mean Visibility**
Units	**%**	
10/06/1987	90	5300
28/06/1987	100	0

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

(a) write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility.

(1)

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

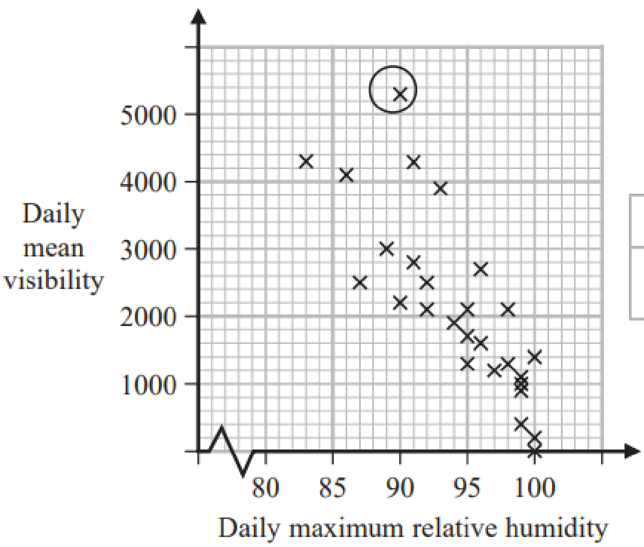

Figure 2

	Q_1	IQR
Daily mean visibility	1100	1600
Daily maximum relative humidity (%)	92	8

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above Q_3 or more than 1.5 times the interquartile range below Q_1.

(b) Show that the point circled on the scatter diagram is an outlier for visibility.

(2)

(c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity.

(1)

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the x–axis.

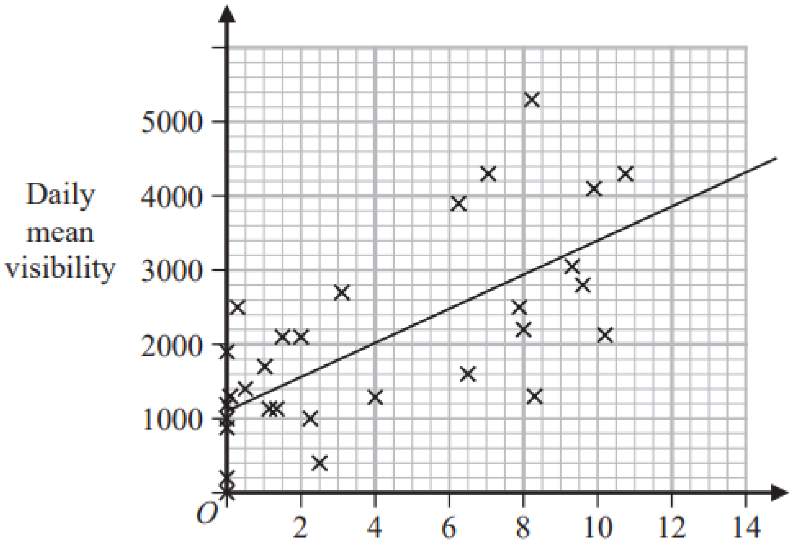

Figure 3

(d) Using your knowledge of the large data set, suggest which variable the x-axis on this scatter diagram represents.

(1)

ANSWER

Question	Scheme	Marks	AOs
2(a)	0 to 500 m	B1	1.2
		(1)	
(b)	$1100+1600+1.5\times1600$ [= 5100]	M1	2.1
	$5300 > 5100$ therefore outlier	A1	1.1b
		(2)	
(c)	As the humidity increases the mean visibility decreases	B1	2.4
		(1)	
(d)	(Hours of) sunshine	B1	2.2b
		(1)	
			(5 marks)

June 2019 question 1

1. A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

(a) Describe how this sample could be taken.

(3)

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s$$

(b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

(1)

(c) Describe one limitation of this regression model.

(1)

ANSWER

Question	Scheme	Marks	AOs
1(a)	Label **each** year group	B1	1.1b
	Use random numbers to select a ...	B1	1.1b
	Simple random sample of 24 Year 12s and 16 Year 13s.	B1	1.1b
		(3)	
(b)	Increase by 2.8 marks	B1	3.4
		(1)	
(c)	e.g. 'the best performance is predicted for the students who never wake up'	B1	3.5b
		(1)	
			(5 marks)

June 2018 Question 1

1. 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)

ANSWER

Qu	Scheme	Marks	AO
1 (a)	Positive (correlation)	B1 (1)	1.2
(b)	Every extra point gives £4.5(0) more on pay (o.e.)	B1 (1)	3.4
(c)	e.g. For points < 11 it would give pay < 0 which is ridiculous	B1 (1)	2.4
		(3 marks)	