1 The table shows some data collected on the temperature, in °C, of a cup of coffee, c, and the time, t in minutes, after which it was made.

t	0	2	4	5	7	11	13	17	25
c	81.9	75.9	70.1	65.1	60.9	51.9	50.8	45.1	39.2

The data is coded using the changes of variable $x = t$ and $y = \log_{10} c$.

The regression line of y on x is found to be $y = 1.89 - 0.0131x$.

a Given that the data can be modelled by an equation of the form $c = ab^t$ where a and b are constants, find the values of a and b. **(3 marks)**

b Give an interpretation of the constant b in this equation. **(1 mark)**

c Explain why this model is not reliable for estimating the temperature of the coffee after an hour. **(1 mark)**

2 The number of bacteria, n thousand per cm^3, in a sample of liquid is measured over a period of time, t, in hours. The data is shown in the table.

t	3.9	5.5	6.8	8.5	10.6	11.5	13.3	14.7	16.5	17.8
n	10.1	13.1	14.6	20.7	27.9	31.5	40	49.9	64.7	75.6

The data is coded using the changes of variable $x = t$ and $y = \log_{10} n$.

The regression line of y on x is found to be $y = 0.7606 + 0.0635x$.

a Given that the data can be modelled by an equation of the form $n = ab^t$ where a and b are constants, find the values of a and b. **(3 marks)**

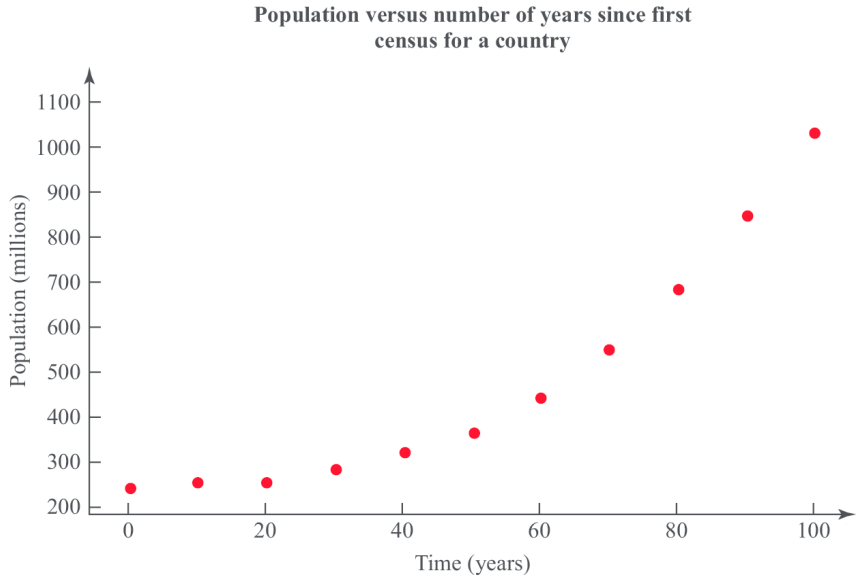
b Give an interpretation of the constant a in this equation. **(1 mark)**

c Explain why this model is not reliable for estimating the number of bacteria after 24 hours. **(1 mark)**

3 The data and scatter diagram show the population, p, in millions, of a country taken t years since their first census.

t	0	10	20	30	40	50	60	70	80	90	100
p	238.4	252.1	251.3	279	318.7	361.1	439.2	548.2	683.3	846.4	1028.7

Figure 1



a Give a reason why the data is coded using the changes of variable $x = t$ and $y = \log_{10} p$. **(1 mark)**

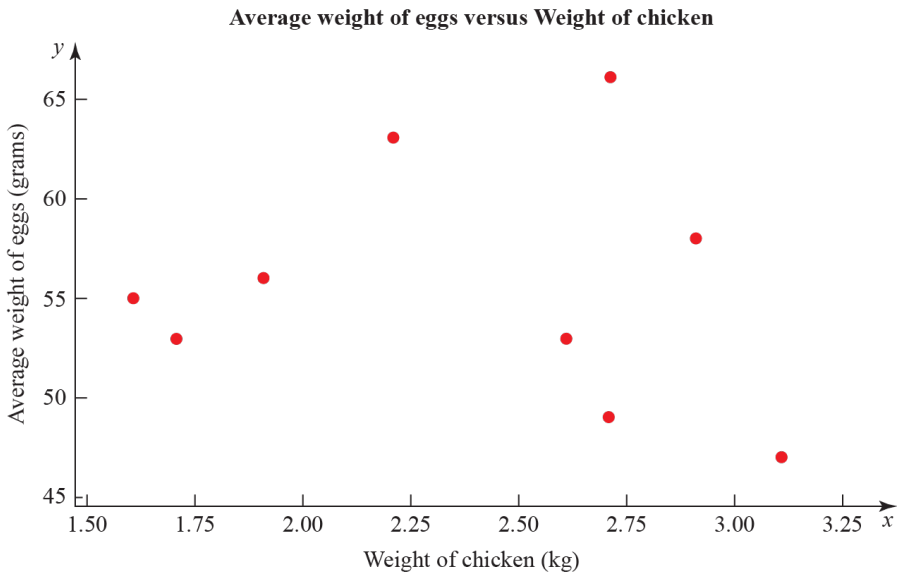
b The product moment correlation coefficient for the coded data is $r = 0.9735$. Comment on r for this model. **(2 marks)**

c With reference to your answer to part **b**, state whether a model in the form $p = ab^t$, where a and b are constants, is a good fit for this data. **(2 marks)**

4 The data and scatter diagram show the weight of chickens, x kilograms, and the average weight, y grams, of eggs laid by a random sample of 10 chickens.

Weight of chickens (kg)	2.9	1.9	1.6	2.7	3.1	2.2	2.7	1.9	1.7	2.6
Average weight of eggs (g)	58	56	55	66	47	63	49	56	53	53

Figure 2



The product moment correlation coefficient for the average weight of eggs and weight of chickens is −0.136.

a Test for evidence of a negative population product moment correlation coefficient at the 2.5% significance level. Interpret this result in context. **(3 marks)**

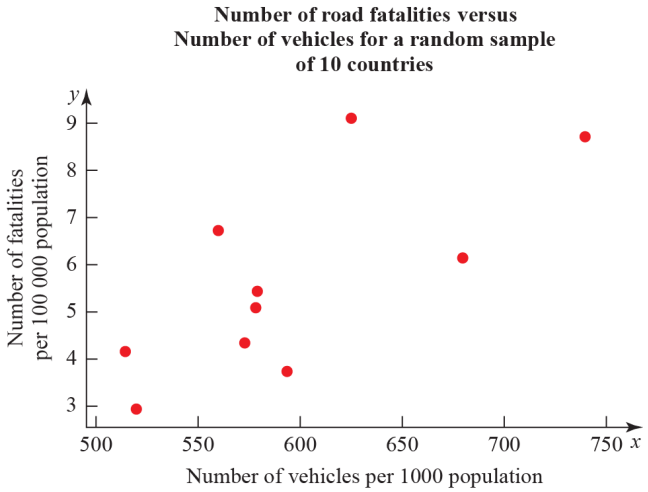
b Explain why even if the population product moment correlation coefficient between two variables is close to zero there may still be a relationship between them. **(2 marks)**

5 A researcher wishes to investigate if there is a positive correlation between the number of vehicles and the number of road fatalities in European countries.

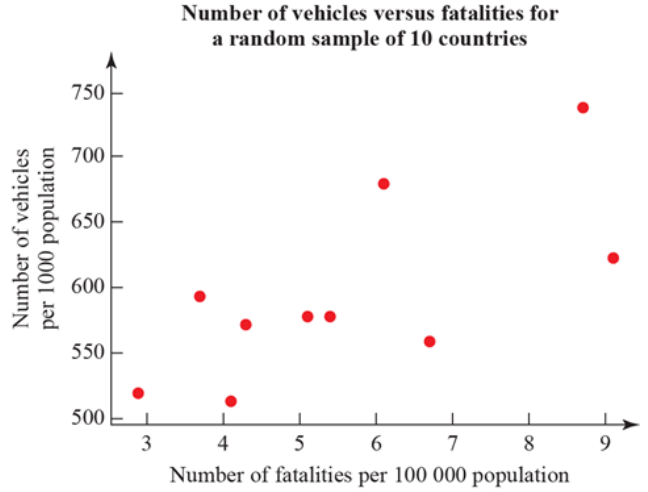
He selects a random sample of 10 European countries and records the number of vehicles, v per 1000 people, and the number of road fatalities, r per 100 000 population, for a particular year. These are shown in the table and scatter diagrams.

Country	v	r
Austria	578	5.4
Belgium	559	6.7
France	578	5.1
Germany	572	4.3
Greece	624	9.1
Ireland	513	4.1
Italy	679	6.1
Luxembourg	739	8.7
Spain	593	3.7
UK	519	2.9

Figure 3



$r = -7.0 + 0.02y$



$v = 460.6 + 24.0r$

a What is the definition of a critical value? **(1 mark)**

b The product moment correlaton coefficient for v and r is 0.714. Use this value to test for positive correlation at the 5% significance level. Interpret your result in context. **(3 marks)**

c The researcher wishes to predict the number of road fatalities for a country with 650 vehicles per 1000 people. Write down the regression model he should use. **(1 mark)**

d State the dependent variable for the regression model in part **c**. **(1 mark)**

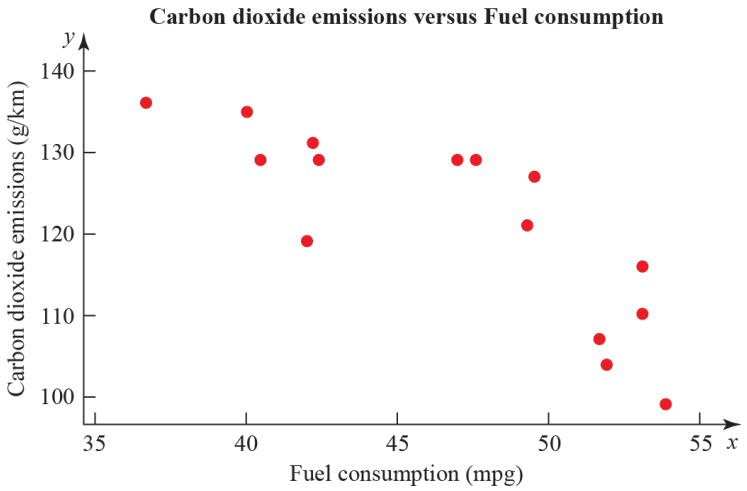
e Monaco has 899 vehicles per 1000 people. Explain why the model stated in **c** is not reliable for estimating the number of road fatalities in Monaco. **(1 mark)**

6 An engineer believes that there is a relationship between the CO_2 emissions and fuel consumption for cars.

A random sample of 40 different car models (old and new) was taken and the CO_2 emission figures, e grams per kilometre, and fuel consumption, f miles per gallon, were recorded.

The engineer calculates the product moment correlation coefficient for the 40 cars and obtains $r = -0.803$.

Figure 4



a State what is measured by the product moment correlation coefficient. **(1 mark)**

b State, with a reason, whether a linear regression model based on these data is reliable or not for a car when the fuel consumption is 60 mpg. **(1 mark)**

c For the linear regression model $e = 198 - 1.71 \times f$ write down the explanatory variable. **(1 mark)**

d State the definition of a hypothesis test. **(1 mark)**

e Test at 1% significance level whether or not the product moment correlation coefficient for CO_2 emissions and fuel consumption is less than zero. State your hypotheses clearly. **(3 marks)**

7 To investigate if there is a correlation between daily mean temperature (°C) and daily mean pressure (hPa) the location Hurn 2015 was randomly selected from:

Camborne 2015	Camborne 1987
Hurn 2015	Hurn 1987
Leuchars 2015	Leuchars 1987
Leeming 2015	Leeming 1987
Heathrow 2015	Heathrow 1987

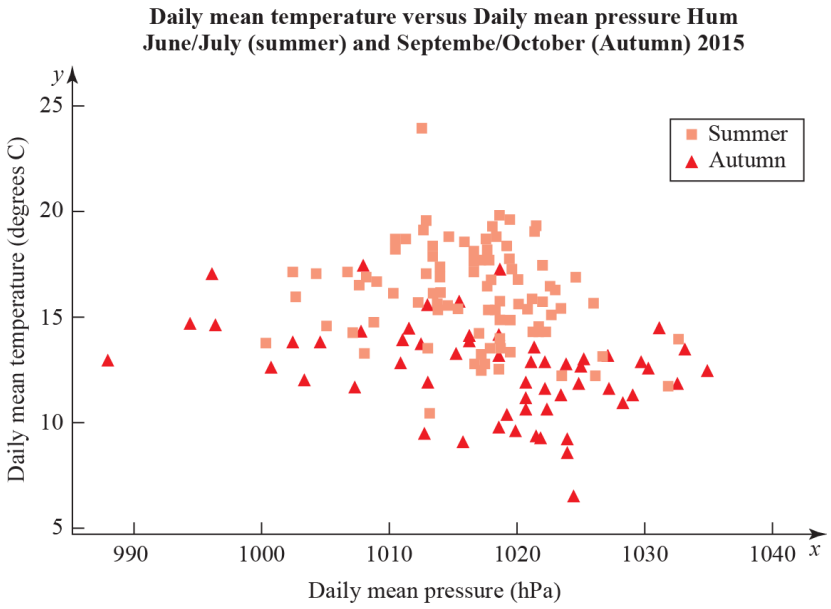
Source: Pearson Edexcel GCE AS and AL Mathematics data set

a State the definition of a test statistic. **(1 mark)**

b The product moment correlation coefficient between daily mean temperature and daily mean pressure for these data is −0.258 with a p-value of 0.001. Use a 5% significance level to test whether or not there is evidence of a correlation between the daily mean temperature and daily mean pressure. **(3 marks)**

c The scatter diagram shows daily mean temperature versus daily mean pressure, by season, for Hurn 2015. Give two interpretations on the split of the data between summer and autumn. **(2 marks)**

Figure 5



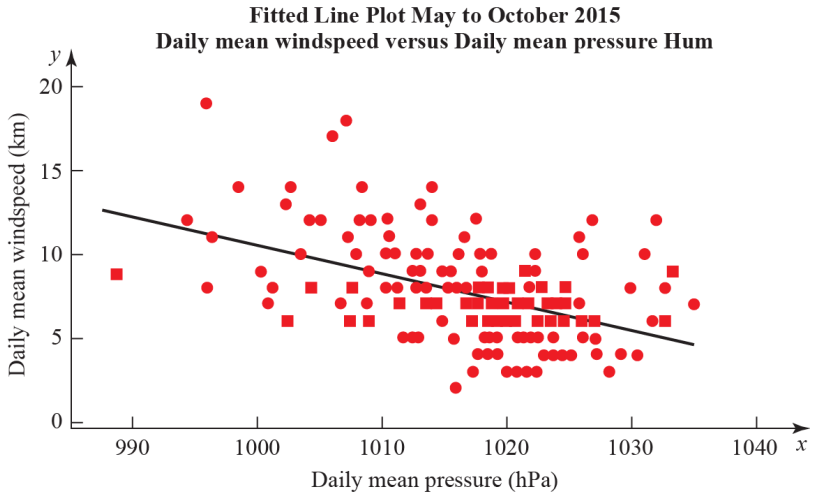
8 To investigate if there is a correlation between daily mean pressure (hPa) and daily mean wind speed (kn) the location Hurn 2015 was randomly selected from:

Camborne 2015	Camborne 1987
Hurn 2015	Hurn 1987
Leuchars 2015	Leuchars 1987
Leeming 2015	Leeming 1987
Heathrow 2015	Heathrow 1987.

Source: Pearson Edexcel GCE AS and AL Mathematics data set

The statistical software output for these data is shown below.

Figure 6



Correlation coefficient
Daily mean winds and Daily mean pressure = −0.477 p-value < 0.001
Regression summary output for daily mean wind speed versus daily mean pressure

	Coefficients	Lower 95%	Upper 95%
Intercept	180.00	133.5424	226.4128
Daily Mean Pressure(hPa) Gradient	−0.1694	−0.21512	−0.12377

a State what is measured by the product moment correlation coefficient. **(1 mark)**

b Comment on the correlation between the two variables. **(1 mark)**

c Give an interpretation of the correlation between the two variables. **(1 mark)**

d Test at 5% significance level whether or not the product moment correlation coefficient for the population is less than zero. State your hypotheses clearly. **(3 marks)**

e Write down the regression model for daily mean wind speed versus daily mean pressure. **(2 marks)**

f Interpret the gradient of the line of regression stated in part **e**. **(1 mark)**

g The regression model (equation of regression) was used to predict the daily mean wind speed of 11.15 knots for a daily mean pressure of 995 hPa. Comment on the accuracy of this prediction. **(1 mark)**