

Year 2 Applied Chapter 1: Regression Correlation Hypothesis Testing
Exam Questions (Total Marks 42)

- 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 $y = \log_{10} c$.

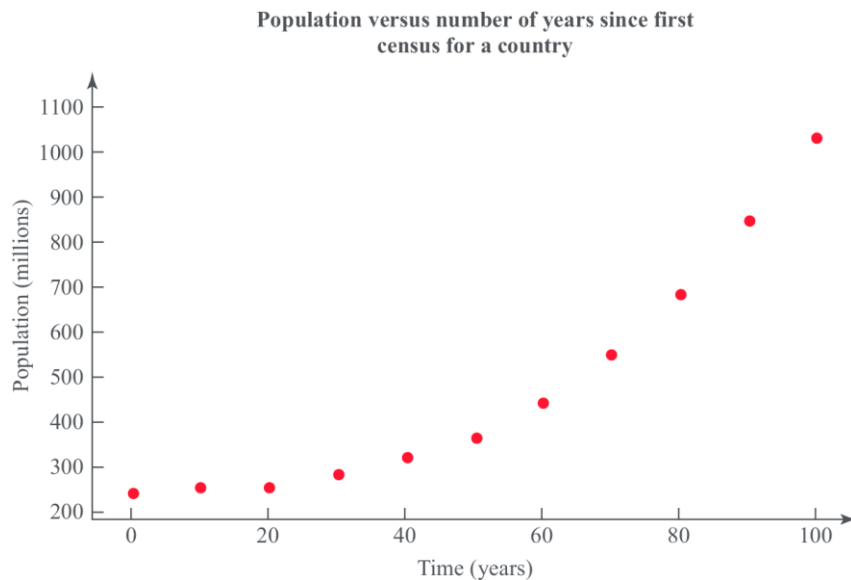
The regression line of y on x is found to be $y = 1.89 - 0.0131x$.

- 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 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



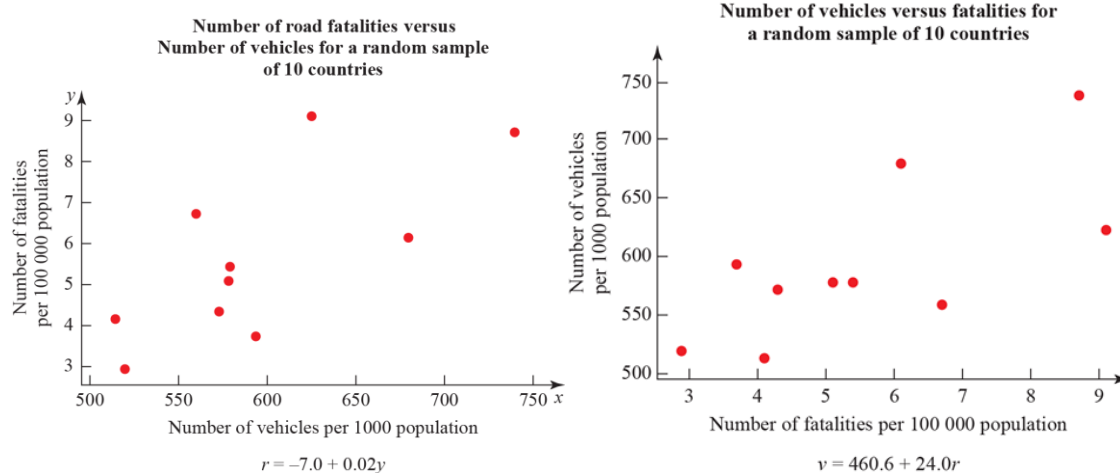
- a Give a reason why the data is coded using the changes of variable $x = t$ and $y = \log_{10} p$. **(1 mark)**
- b The product moment correlation coefficient for the coded data is $r = 0.9735$. Comment on r for this model. **(2 marks)**
- c With reference to your answer to part b, state whether a model in the form $p = ab^t$, where a and b are constants, is a good fit for this data. **(2 marks)**

- 3) 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



- What is the definition of a critical value? (1 mark)
- The product moment correlation 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)
- 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)
- State the dependent variable for the regression model in part c. (1 mark)
- 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)

- 4) To investigate if there is a correlation between daily mean temperature ($^{\circ}\text{C}$) and daily mean pressure (hPa) the location Hurn 2015 was randomly selected from:

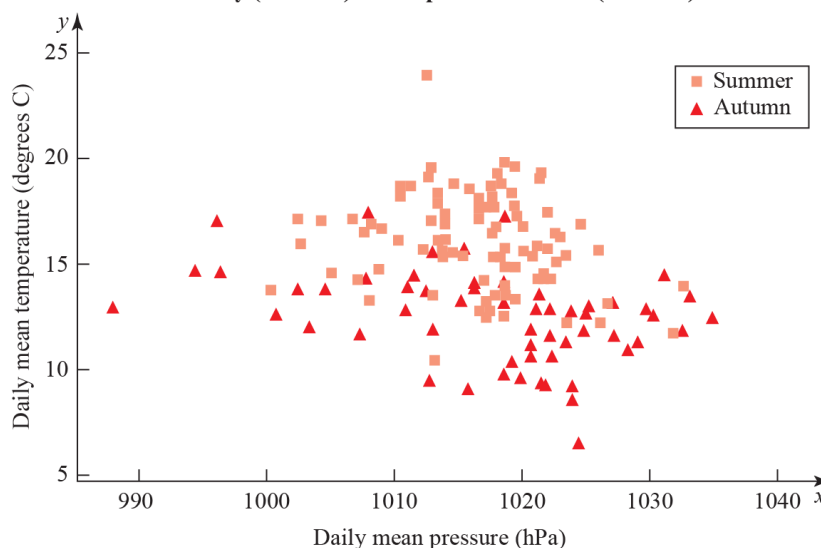
Camborne 2015	Camborne 1987
Hurn 2015	Hurn 1987
Leuchars 2015	Leuchars 1987
Leeming 2015	Leeming 1987
Heathrow 2015	Heathrow 1987

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

**Daily mean temperature versus Daily mean pressure Hurn
June/July (summer) and September/October (Autumn) 2015**



5)

The temperature and the rainfall on 15 days is recorded.

(a) Suggest a null and alternative hypothesis for a two tailed test to investigate whether there is a correlation between temperature and rainfall. **(2)**

The product moment correlation coefficient is calculated to be $r = 0.37$.

(b) Test your hypotheses at the 10% significance level. **(2)**

6)

The temperature and the number of hours of sunshine on 12 days is recorded.
The product moment correlation coefficient is calculated to be $r = 0.636$.

Stating your hypotheses clearly test, at the 1% significance level, whether there is a positive correlation between temperature and hours of sunshine

(4)

7)

The temperature of water ($w^{\circ}\text{C}$) in a kettle t minutes after it was boiled is recorded in the table below

t	1	3	6	9	11	15	20
w	92	76	65	50	46	33	25

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

The regression line of x on y is found to be $y = 1.99 - 0.031x$

(a) Given that the data can be modelled by an equation in the form $w = ab^t$ where a and b are constants. Find the values of a and b to 3 significant figures. (3)

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

(c) Explain why this model is not reliable for calculating the temperature after 1 hour. (1)

8)

The variables x and y are recorded and the results are shown in the table below

x	1	2	3	4	5	6	7	8	9	10
y	429	754	871	1119	2478	2653	3050	3279	5470	7439

The data is coded using the changes of variable $r = x$ and $s = \log_{10} y$

(a) The product moment correlation coefficient for the coded data is 0.98
Comment on r for this model and therefore justify the use of a model in the form $y = ab^x$ here a and b are constants **(3)**

The regression line of r on s is found to be $s = 2.60 + 0.127r$

(b) Find the values of a and b , give your answers to 3 significant figures **(3)**