Lesson 6

A-Level Applied Mathematics : Statistics : Year 2

Correlation II

6.1 Revision

Any solution based entirely on graphical
or numerical methods is not acceptable

Marks Available: 50

Question 1

An ice-cream seller believes that there is a positive correlation between the amount of sunshine and sales of ice cream. He collects data on six days during June 2015 at his pitch in Camborne:

Sunshine (hours)	4.2	7.9	13.8	8.7	6.2	0.7
Ice-Cream sales (£100s)	7.0	8.3	12.4	8.1	7.9	6.2

(a) Use your statistics calculator to find the product moment correlation coefficient, the PMCC, for these data

[2 marks]

(b) Carry out a hypothesis test to determine, at the 5% level, if there is significant evidence in support of the ice-cream seller's belief.
State your hypotheses clearly.

[4 marks]

Question 2

It is found that when a particular set of data is coded using

$$Y = \log y \qquad \text{and} \qquad X = \log x$$

a straight line graph is obtained when Y is plotted against X
State two important facts that can be deduced about the relationship between y and x.

[2 marks]

Question 3

A-Level Examination Question from 17th May 2013, paper S1, Q1 (Edexcel)

A meteorologist believes that there is a relationship between the height above sea level, h m, and the air temperature, t °C.

Data is collected at the same time from 9 different places on the same mountain.

The data is summarised in the table below.

h	1400	1100	260	840	900	550	1230	100	770
t	3	10	20	9	10	13	5	24	16

(a) Use your statistics calculator to determine the product moment correlation coefficient for this data, and also the equation of the straight line of best fit which will be of the form $t = a + b\,h$

Give values in your answers to 3 significant figures.

[3 marks]

(b) State whether or not your part (a) answer supports the use of a regression equation to predict the air temperature at different heights on this mountain. Give a reason for your answer.

[2 marks]

(c) Interpret the value of b from your line of best fit equation, $t = a + b\,h$

[1 mark]

(d) Estimate the difference in air temperature between a height of 500 m and a height of 1 000 m

[3 marks]

Question 4

The number of atoms of a radioactive substance, n, is measured at various times, t minutes after the start of an experiment. The table below shows the data.

Time, t	1	2	4	5	7
Atoms, n	231	41	17	7	2
$\log n$					

The data is coded using

$$y = \log n \quad \text{and} \quad x = t$$

(a) Complete the table showing the values of $\log n$
Work to three decimal places.

[2 marks]

(b) Calculate the product moment correlation coefficient for the coded data.

[2 marks]

(c) With reference to your answer to part (b), state whether an exponential model is a good fit.

[2 marks]

The equation of the regression line of y on x is found to be

$$y = 2.487 - 0.320\,x$$

(d) Find an expression for n in terms of t, giving your answer in the form

$$n = a\,b^t$$

where a and b are constants to be found.

[3 marks]

Question 5

An engineer wants to see if there is a connection between the speed of a car and its fuel consumption.

The results from a random selection of 10 speeds of the same car are as follows;

Speed (mph)	75	30	52	98	26	88	106	49	65	76
Fuel consumption (mpg)	30	66	64	19	40	42	17	57	50	47

(a) State suitable null and alternative hypotheses

[2 marks]

The product moment correlation coefficient was calculated to be – 0.749

(b) Carry out the hypothesis test at the 1% significance level.

[4 marks]

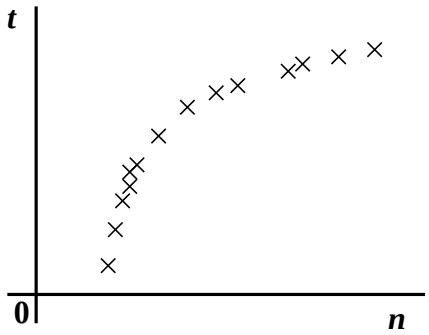
(c) Would your conclusion be the same at the 5% significance level ?
Explain your answer.

[4 marks]

Question 6

The time in milliseconds, t, needed for a computer algorithm to determine whether a number, n, is prime is recorded for different values of n.

A scatter graph of t against n is drawn.



(a) Explain why a model of the form $t = a + b\,n$ is unlikely to fit these data.

[1 mark]

(b) The data are coded using the changes of variable

$$y = \log t \qquad \text{and} \qquad x = \log n$$

The regression line of y on x is found to be

$$y = -0.301 + 0.6x$$

Find an equation for t in terms of n giving your answer in the form

$$t = a\,n^k$$

where a and k are constants to be found

[4 marks]

(c) How long will it take to test the number 990486497 ?

Give your answer in minutes and seconds, correct to the nearest second.

[2 marks]

Question 7

Officials are testing athletes for doping at a sporting event. They model the concentration of a particular drug in an athlete's bloodstream using the equation

$$D = 6e^{-0.1t}$$

where D is the concentration of the drug in mg/l and t is the time in hours since the athlete took the drug.

(a) Interpret the meaning of the constant 6 in this model.

[1 mark]

(b) Find the concentration of the drug in the bloodstream after 2 hours.

[1 mark]

(c) It is impossible to detect this drug in the bloodstream if the concentration is lower than 3 mg/l. Find the time in hours and minutes, to the nearest minute, when the drug first becomes undetectable.

[2 marks]

(d) Equations of the form (the best ***e*** curve)

$$y = ae^{bx}$$

can be detected by a coding that will transform them into a straight line. Analyse this curve by using natural logarithms, to determine what the coding should be to test for this type of curve.

[3 marks]