Correlation and Regression

1 A survey of British towns recorded the number of serious road accidents in a week (x) in each town, together with the number of fast food restaurants (y). The data showed a strong positive correlation. Katie states that this shows that building more fast food restaurants in her town will cause more serious road accidents. Explain whether the data supports Katie's statement.

2 The following table shows the mean CO_2 concentration in the atmosphere, c (ppm), and the increase in average temperature compared to the 30-year period 1951–1980, t (°C).

Year	2015	2013	2011	2009	2007	2005	2003	2001	1999	1997	1995	1994
c (ppm)	401	397	392	387	384	381	376	371	368	363	361	357
t (°C)	0.86	0.65	0.59	0.64	0.65	0.68	0.61	0.54	0.41	0.47	0.45	0.24

Source: Earth System Research Laboratory (CO_2 data); GISS Surface Temperature Analysis, NASA (temperature data)

a Draw a scatter diagram to represent this data.

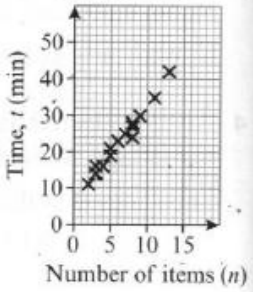
b Describe the correlation between c and t.

c Interpret your answer to part **b**.

3 The table below shows the packing times for a particular employee for a random sample of orders in a mail order company.

Number of items (n)	2	3	3	4	5	5	6	7	8	8	8	9	11	13
Time (t min)	11	14	16	16	19	21	23	25	24	27	28	30	35	42

A scatter diagram was drawn to represent the data.



a Describe the correlation between number of items packed and time taken. **(1 mark)**

The equation of the regression line of t on n is $t = 6.3 + 2.64n$.

b Give an interpretation of the value 2.64. **(1 mark)**

4 Energy consumption is claimed to be a good predictor of Gross National Product. An economist recorded the energy consumption (x) and the Gross National Product (y) for eight countries. The data is shown in the table.

Energy consumption (x)	3.4	7.7	12.0	75	58	67	113	131
Gross National Product (y)	55	240	390	1100	1390	1330	1400	1900

The equation of the regression line of y on x is $y = 225 + 12.9x$.

The economist uses this regression equation to estimate the energy consumption of a country with a Gross National Product of 3500.

Give two reasons why this may not be a valid estimate. **(2 marks)**

5 The table shows average monthly temperature, t (°C), and the number of pairs of gloves, g, a shop sells each month.

t (°C)	6	6	50	10	13	16	18	19	16	12	9	7
g	81	58	50	42	19	21	4	2	20	33	58	65

The following statistics were calculated for the data on temperature:
mean = 15.2, standard deviation = 11.4

An outlier is an observation which lies ±2 standard deviations from the mean.

a Show that $t = 50$ is an outlier. **(1 mark)**

b Give a reason whether or not this outlier should be omitted from the data. **(1 mark)**

The equation of the regression line of t on g for the remaining data is $t = 18.4 - 0.18g$.

c Give an interpretation of the value -0.18 in this regression equation. **(1 mark)**

6 James placed different masses (m) on a spring and measured the resulting length of the spring (s) in centimetres. The smallest mass was 20 g and the largest mass was 100 g.

He found the equation of the regression line of s on m to be $s = 44 + 0.2m$.

a Interpret the values 44 and 0.2 in this context. **(2 marks)**

b Explain why it would not be sensible to use the regression equation to work out:

i the value of s when $m = 150$ **ii** the value of m when $s = 60$. **(2 marks)**

7 A student is investigating the relationship between the price (y pence) of 100 g of chocolate and the percentage (x%) of cocoa solids in the chocolate.

The data obtained is shown in the table.

Chocolate brand	x (% cocoa)	y (pence)
A	10	35
B	20	55
C	30	40
D	35	100
E	40	60
F	50	90
G	60	110
H	70	130

a Draw a scatter diagram to represent this data. **(2 marks)**

The equation of the regression line of y on x is $y = 17.0 + 1.54x$.

b Draw the regression line on your diagram. **(2 marks)**

The student believes that one brand of chocolate is overpriced and uses the regression line to suggest a fair price for this brand.

c Suggest, with a reason, which brand is overpriced. **(1 mark)**

d Comment on the validity of the student's method for suggesting a fair price. **(1 mark)**