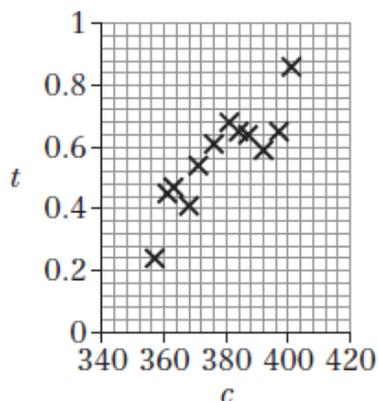


Correlation, Mixed Exercise 4

- 1 The data shows that the number of serious road accidents in a week strongly correlates with the number of fast food restaurants. However, it does not show whether the relationship is causal. Both variables could correlate with a third variable, e.g. the number of roads coming into a town.

2 a



- b There is strong positive correlation.
- c As mean CO₂ concentration in the atmosphere increased, mean temperatures also increased.
- 3 a There is strong positive correlation.
- b If the number of items increases by 1, the time taken increases by approximately 2.64 minutes.
- 4 The answer is likely to be unreliable as it involves extrapolation. 3500 is well outside the limits of the data set used.

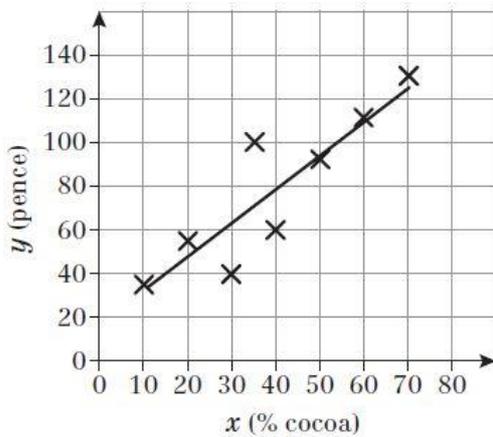
The economist used the regression line of y on x . Estimating energy consumption (x) from Gross National Product (y) would need the regression line of x on y .

5 a $15.2 + 2 \times 11.4 = 38$

As $50 > 38$, $t = 50$ °C is an outlier.

- b The outlier should be omitted, as it is very unlikely that the average temperature was 50 °C in a climate where people need to buy gloves, and so this data point is likely an anomaly.
- c The equation of the regression line of g on t is $g = 99.6 - 5.2t$.
- This means that for every increase in temperature of 1°C, the shop sells 5.2 fewer pairs of pairs of gloves.
- 6 a 44 cm is the length of the spring with no mass attached. If a mass of 1 g is attached, the spring will increase in length by approximately 0.2 cm.
- b i $m = 150$ is outside the range of the data (extrapolation) so is less likely to be accurate.
- ii This particular regression equation should only be used to predict a value of s given m . To predict a value of m given s , you should use the regression equation of m on s .

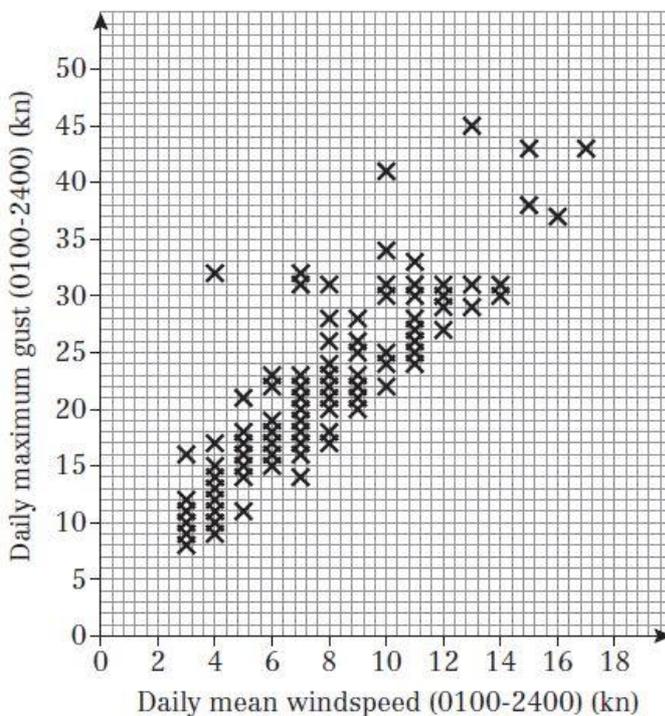
7 a and b



- c Brand D is overpriced, since it's price is much more than you would expect (the data point is far above the regression line)..
- d The regression equation should be used to predict a value for y given x, i.e. the price given the percentage of cocoa solids. So the student's method is a valid one.

Large data set

1 a



- b There is moderate positive correlation
- c The relationship is causal, as the maximum gust is related to the mean windspeed. A higher mean windspeed is likely to result in a higher maximum gust.
- d i $g = 4.97 + 2.15 \times 0.5$
 $= 6.045$

d ii $g = 4.97 + 2.15 \times 5$
 $= 15.72 \text{ kn}$

iii $g = 4.97 + 2.15 \times 12$
 $= 30.77 \text{ kn}$

iv $g = 4.97 + 2.15 \times 40$
 $= 90.97 \text{ kn}$

e Parts **ii** and **iii** are within the range of the data (interpolation), so are more likely to be accurate. Parts **i** and **iv** are outside the range of the data (extrapolation), so are less likely to be accurate.

f Using the 'SLOPE' and 'INTERCEPT' function in a spreadsheet:

$$w = 0.05 + 0.35 g$$

$$w = 0.05 + 0.35 \times 30$$

$$= 10.55 \text{ kn}$$

2 a Using the 'SLOPE' and 'INTERCEPT' function in a spreadsheet:

$$s = 15.10 - 1.80 c \text{ (values taken to 2 d.p.)}$$

The missing total sunshine data values are approximately:

1.85, 8.51, 6.27, 11.58, 2.97, 7.09, 12.18, 14.42, 9.48, 6.27, 2.08, 3.87, 7.16, 3.12, 3.72, 3.87, 0.95

b The relationship is likely to be causal: greater cloud cover is likely to result in less sunshine.