Correlation and Regression

E) 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)**

a)   Outlier if $> \bar{t} + 2\,s.d_t$

$15.2 + 2 \times 11.4 = 38$

$50 > 38$ ∴ $t = 50$ is an outlier

b)   50°c would not happen in a place with a temperate climate in the other 11 months

c)   The drop in temperature in °C for every extra pair of gloves sold.

The table shows the daily total rainfall, r mm, and daily total hours of sunshine, s, in Leuchars for a random sample of 11 days in August 1987, from the large data set.

r	0	6.8	0.9	4.8	0	21.7	1.7	4.9	0.1	2.2	0.1
s	8.4	4.9	10.2	4.5	3.3	3.9	5.4	1.8	9.7	1	4.6

© Crown Copyright Met Office

The median and quartiles for the rainfall data are: $Q_1 = 0.1$ $Q_2 = 1.7$ $Q_3 = 4.85$

An outlier is defined as a value which lies either 1.5 × the interquartile range above the upper quartile or 1.5 × the interquartile range below the lower quartile.

a Show that $r = 21.7$ is an outlier. (1 mark)

b Give a reason why you might:
   i include   ii exclude this day's readings. (2 marks)

c Exclude this day's readings and draw a scatter diagram to represent the data for the remaining ten days. (3 marks)

d Describe the correlation between rainfall and hours of sunshine. (1 mark)

e Do you think there is a causal relationship between the amount of rain and the hours of sunshine on a particular day? Explain your reasoning. (1 mark)
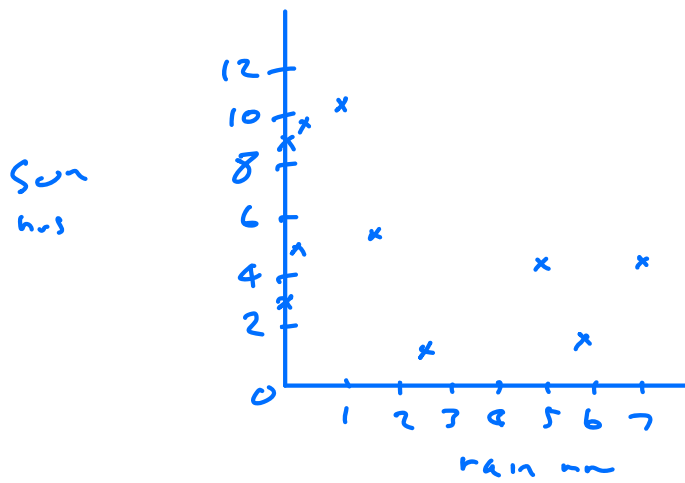
a) outlier if > $Q_3 + 1.5 \times IQ$

$= 4.85 + 1.5 \times (4.85 - 0.1)$

$= 11.975$

$r = 21 > 11.975$ so $r = 21$ is an outlier

b) include   –   could be a thunderstorm that brought unusual heavy rain

exclude   –   extremely unusual for August so could be a recording error

Sun
hrs

(scatter graph with y-axis "Sun hrs" labeled 0, 2, 4, 6, 8, 10, 12 and x-axis "rain mm" labeled 1, 2, 3, 4, 5, 6, 7)

weak negative correlation

No causal relationship between rain and sun

Perhaps cloud cover and sun, but can have cloud without rain