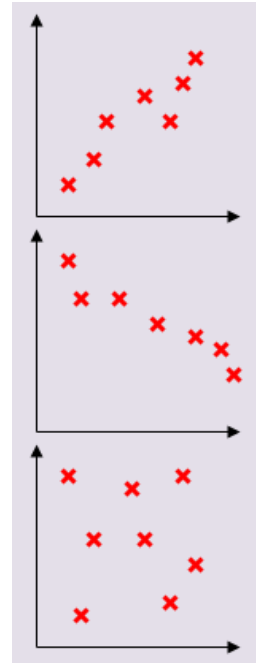


4A Correlation & Drawing Scatter Diagrams

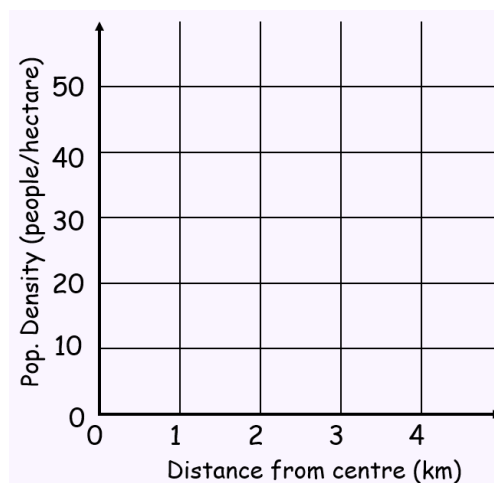


1. In the study of a city, the population density, in people/hectare, and the distance from the city centre, in km, was investigated by choosing sample areas. The results are as follows:

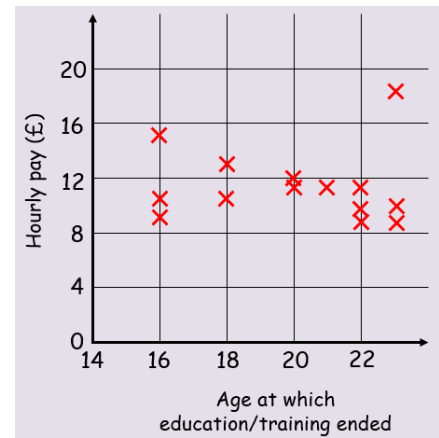
Area	A	B	C	D	E
Distance	0.6	3.8	2.4	3.0	2.0
Pop. Density	50	22	14	20	33

Area	F	G	H	I	J
Distance	1.5	1.8	3.4	4.0	0.9
Pop. Density	47	25	8	16	38

Plot a scatter graph and describe the correlation. Interpret what the correlation means.



2. Hideko was interested to see if there was a relationship between what people earn and the age at which they left education or training. She asked 14 friends to fill in an anonymous questionnaire and recorded her results in a scatter diagram.
- a) Describe the type of correlation shown



Hideko says that her data supports the conclusion that more education causes people to earn a lower hourly rate of pay

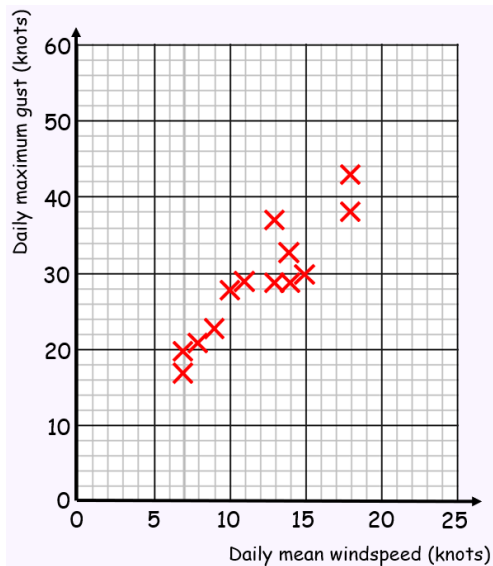
- b) Give one reason why Hideko's conclusion might not be valid

4B Linear Regression - Lines of Best Fit

1. From the large data set, the daily mean windspeed, w knots, and the daily maximum gust, g knots, were recorded for the first 15 days in May in Camborne in 2015.

The data was plotted on a scatter diagram:

- a) Describe the correlation between daily mean windspeed and daily maximum gust



w	g
14	33
13	37
13	29
9	23
18	43
18	38
7	17
15	30
10	28
14	29
11	29
9	23
8	21
10	28
7	20

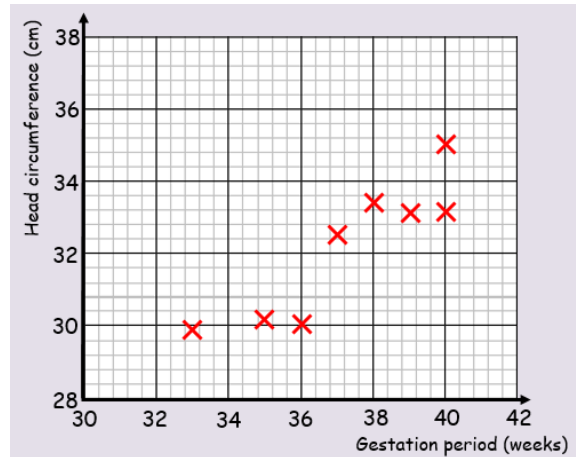
The equation of the regression line of g on w for this data is:

$$g = 7.23 + 1.82w$$

- b) Give an interpretation of the value of the gradient of this regression line
- c) Justify the use of a linear regression line in this case.

2. The head circumference, y cm, and gestation period, x weeks, for a random sample of newborn babies were recorded, and the scatter graph shows the results.

Gestation period, x (weeks)	Head circumference, y (cm)
36	30.0
40	35.0
33	29.8
37	32.5
40	33.2
39	32.1
35	30.9
38	33.6



The equation of the regression line of y on x is:

$$y = 8.91 + 0.624x$$

- a) This equation is used to estimate the head circumference of a baby born after 39 weeks and a baby born after 30 weeks. Comment on the reliability of these estimates.

A nurse wants to estimate the gestation period for a baby born with a head circumference of 31.6cm.

- b) Explain why the regression equation above is not suitable for this estimate