

## 6.2 The line of best fit

- 1** Have you ever wondered why it took so long to find Cinderella? Well the answer is simple: shoe sizes are not internationally set.

The table shows statistics about the foot length along with the size worn in Europe and in US by 15 different people.

Foot length (in cm)	European Size	US Size
21.2	33	4
21.3	34	4.5
21.7	34	4.5
22.6	36	6
23.2	36.5	6.5
23.3	36	6
23.4	37	7
24	38	7.5
24.2	38	7.5
24.3	38.5	8
24.6	39.5	9
25.2	39	8.5
25.3	39	8.5
26.8	42	10.5
26.8	43	11

- a** Analyse the correlation between the foot length and the European size. For this you should:
- i** Plot a scatter diagram for this data
  - ii** Calculate the Pearson product moment correlation coefficient for this data.
  - iii** Describe the correlation
  - iv** Write down the equation of the regression line of  $y$  on  $x$ .
  - v** Find the mean of both variables.
  - vi** Plot and label the mean point on the scatter
  - vii** Draw the line of best fit
  - viii** Hence predict the length of your feet from your shoe size
- b** Repeat **i** to **vii**. For European size and the US size
- viii** Explain why, despite the fact that there is a one-to-one relation between European size and the US size data, the Pearson product moment correlation coefficient is not 1.

- 2 Cinderella and Prince Charming lived happy ever after... What about reality? Below are the annual percentage risk of break up, by year of relationship.

Year(s) together	1	2	3	4	5	8	10	15	20	22	25	28	29	36	40
Annual rate for married couples as a %	4	4	3.5	3.6	3.4	3.2	3.1	3	2	2.2	2	1.8	1.7	1.5	1.2
Annual rate for non-married couples	70	32.5	25	22	20	16	14	12	11	10.8	11	11.3	11.5	13	14.5

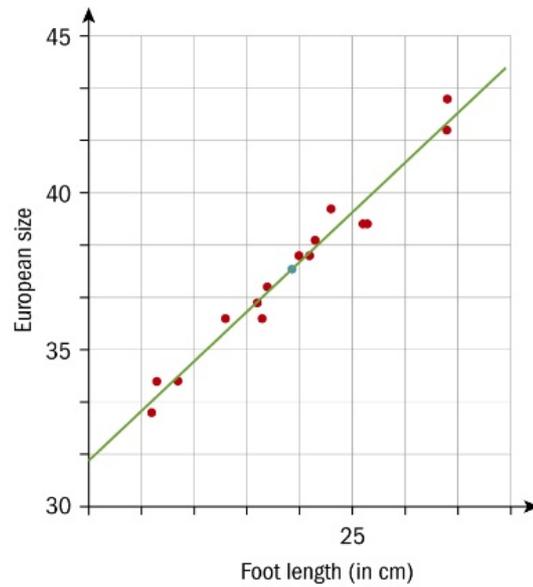
- a Analyse the correlation between the rate of annual break up and the number of years together for married couples. For this you should:
- Plot a scatter diagram for this data
  - Calculate the Pearson product moment correlation coefficient for this data.
  - Describe the correlation
  - Write down the equation of the regression line of  $y$  on  $x$ .
  - Plot and label the mean point on the scatter
  - Draw the line of best fit
  - Predict the rate of chance of breaking up after 30 years
  - After how many years in the relationship there is 0% chance of break-up.
- b Analyse the correlation between the rate of annual break up and the number of years together for unmarried couples. For this you should repeat **i** to **vii**.
- viii Explain how using a piecewise function would be better to map the correlation for non-married couples.
- 3 Tyler Vigen in Spurious correlations analysed the correlation between the consumption of margarine per capita and the divorce rate in Maine. Use the data to:

Margarine consumed (in lbs)	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7
Divorce rate in Main( in %)	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1

- Plot a scatter diagram for this data
- Calculate the Pearson product moment correlation coefficient for this data.
- Describe the correlation
- Write down the equation of the regression line of  $y$  on  $x$ .
- Find the mean quantity of margarine consumed
- Find the mean of the percentage of divorce
- Plot and label the mean point on the scatter
- Draw the line of best fit
- Discuss whether you can predict the percentage of divorce of a couple who eats 9lbs of margarine.

## Answers

1 a i, vi, vii

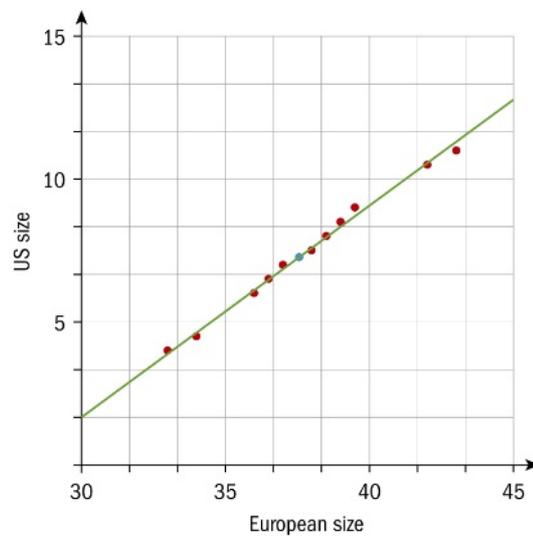
ii  $r=98.37\%$ 

iii very strong

iv  $y = 1.6x - 0.19$ 

v Foot length 23.86, size= 37.57

b i, vi, vii.

ii  $r=99.72\%$ 

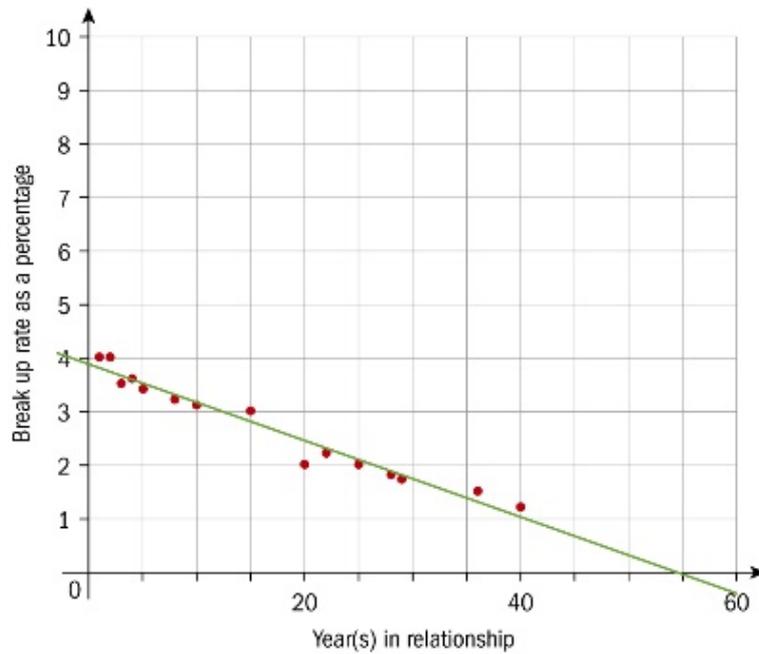
iii very strong

iv  $y = 0.73x - 20.5$ 

v European size= 37.57, US size= 7.27

viii Although it is a one to one relationship it is not a perfectly linear relationship.

## 2 a i, vi



ii -0.98.

iii Strong negative linear correlation

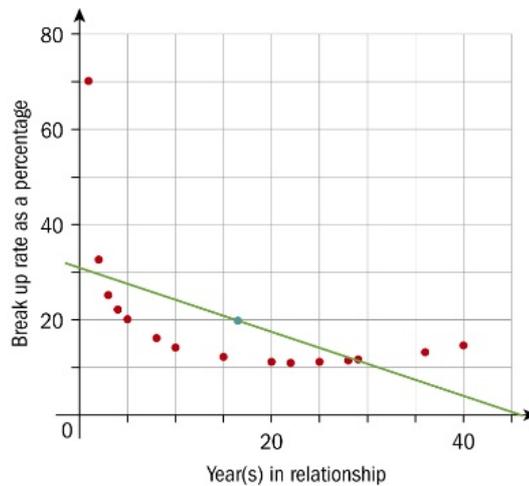
iv  $y = -0.07x + 3.9$ 

v mean is the point (16.53, 2.68)

vi 1.7%

vii 54 years

## b i, vi



ii -0.671

iii Moderate negative correlation

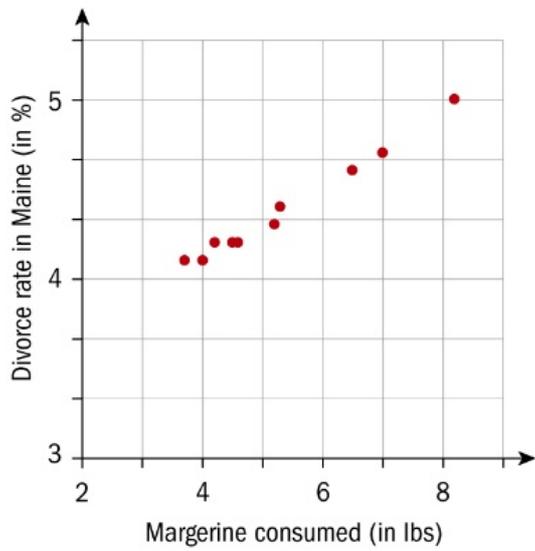
iv  $y = -0.67x + 30.7$ 

v mean point: (16.53, 19.64)

vii 10.6%

viii this mapping seems to be made of 3 parts: a dramatic drop for the first three years, a slow drop till 15 years and then the percentage of break up rise again slightly.

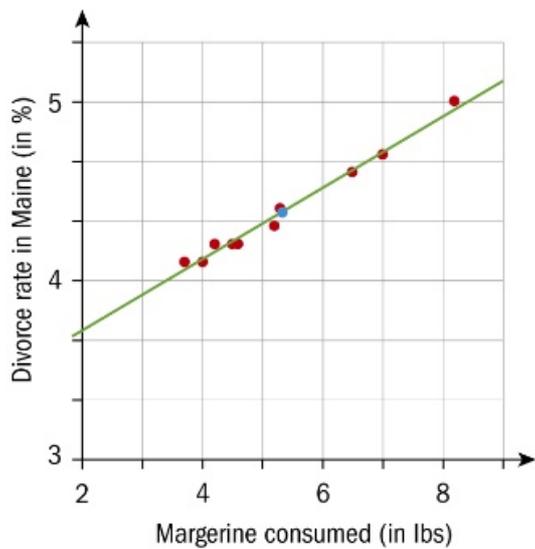
3 a



**b**  $r = 99.26\%$       **c** very strong correlation      **d**  $y = 0.2x + 3.3$

**e** 5.32%      **f** 4.38%

**g, h**



**i** According to the graph if  $x = 9$ ,  $y = 5.121\%$ , this is however a spurious correlation. Those two variables are clearly independent and the correlation cause is due to another factor or mere chance.