

## SCATTER DIAGRAMS

### CORRELATION

If we looked at the exam results of a group of students in different subjects we may not be surprised to see that there seemed to be some link between each student's results in, say, physics and maths. The student who did well in maths also did well in physics; those who did badly in maths also had poor results in physics. If this is the case we say that there is **correlation** between performance in physics and maths.

### Scatter Diagrams

A scatter diagram is a graph which helps us to see how much correlation there is between two variables which we suspect maybe correlated.

#### Example 1

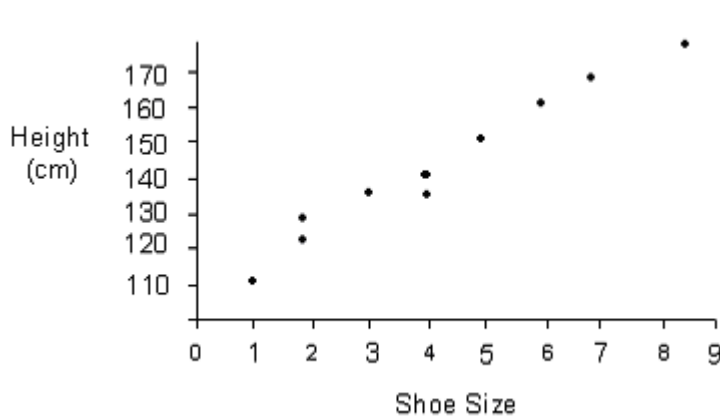
The table below gives us information about the heights and shoe sizes of ten children. From what we know of children we should expect height and shoe size to be correlated. We plot a scatter diagram to check this.

**Table 1**

The table gives the height in cm and shoe size of 10 children.

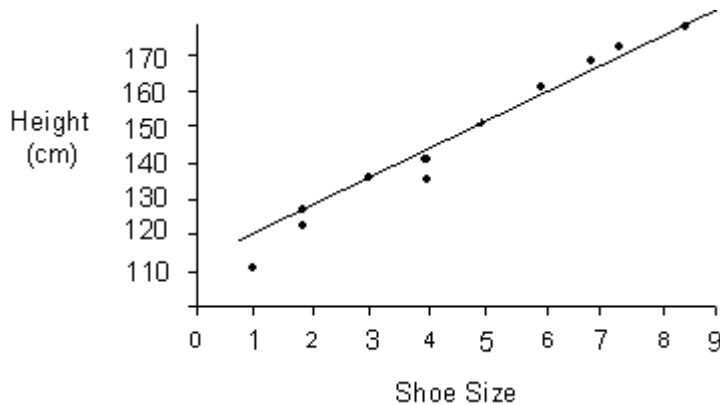
Child	A	B	C	D	E	F	G	H	I	J
Height (cm)	130	135	135	155	150	160	120	170	140	125
Shoe size	2	4	3	6	5	7	1	9	4	2

#### Diagram 1



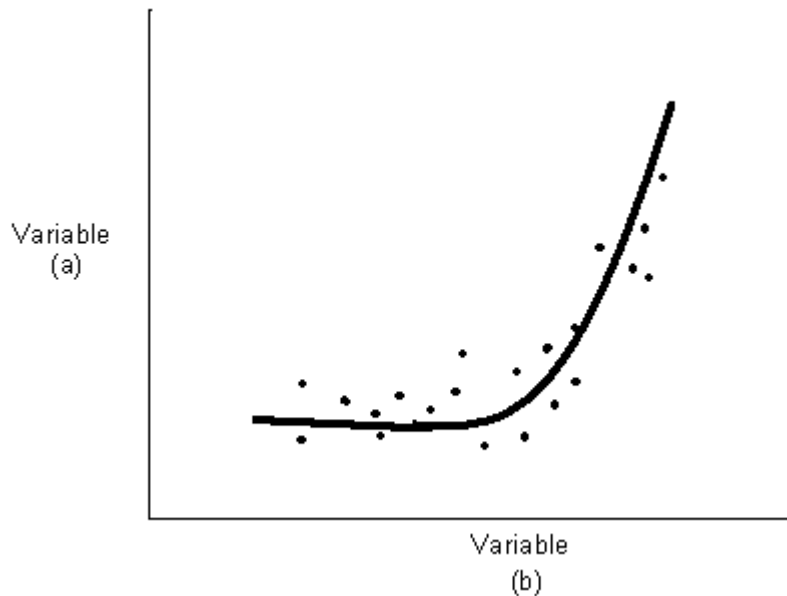
We can see that the points on the diagram are not evenly scattered. The points seem to nearly lie on a line. The line is drawn to be as near as possible to **all** the points. This shows there is a link between the 'height' and 'shoe size'. The line that is drawn is called **the line of best fit or the regression line**. In this example the scatter diagram suggests that the relationship between the shoe size and height is represented by a straight line, it is therefore called a **linear relationship**. Other examples might give a scatter diagram which suggests a curve but these are not included in this pack.

**Diagram 2**



**Diagram 3**

A different scatter diagram which suggests that a curve would best fit the data.



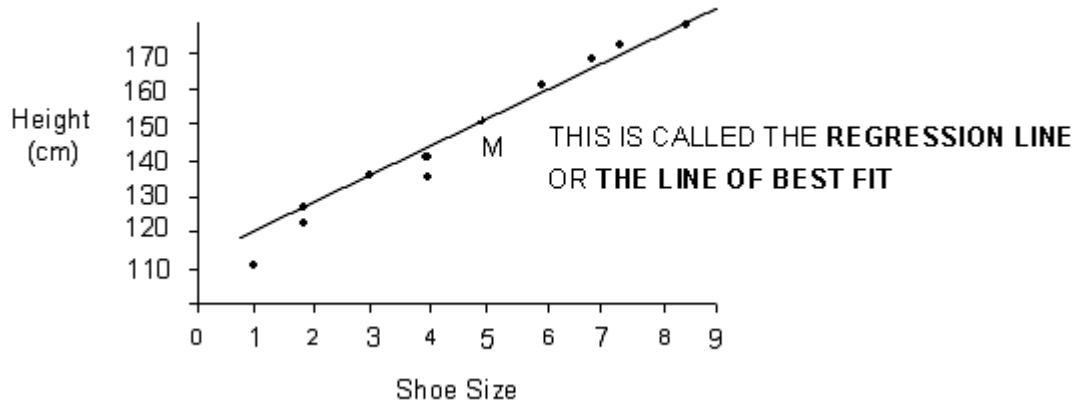
### The Regression Line

To draw the regression line for **example 1**:

1. Plot the scatter diagram as in diagram 1 above.
2. Work out the mean height of the set of 10 children from the data in the table and the mean shoe size.  
 Mean height = 142  
 Mean shoe size = 4.3
3. Plot the point M on the scatter diagram to show the mean height and shoe size.
4. Draw, using a ruler, a line passing through M which gets close as possible to as many points on the scatter diagram as possible. This is the line of best fit.

See Diagram 4.

Diagram 4



### Direct and Indirect Correlation

Example 1 was an example of direct or positive correlation. High values of shoe size went with high values of height and low values went with low values. Other examples may show indirect or negative correlation. Here are some examples of pairs of variables that would probably be negatively correlated.

The number of people on a certain beach one day and the number of millimetres of rain that fell that day.

The price of TV sets and the number sold.

Think of two more examples yourself.

To Re-Cap

Scatter diagrams help us to see how much correlation there is between two variables and also what type of correlation it is. Look at the scatter diagrams on the next page.

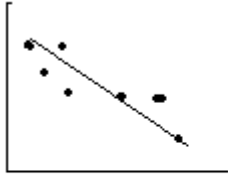


Diagram 5

NEGATIVE CORRELATION  
OR HIGH CORRELATION

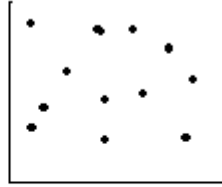


Diagram 6

NO CORRELATION



Diagram 7

PERFECT POSITIVE  
CORRELATION

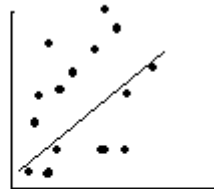


Diagram 8

SOME POSITIVE  
CORRELATION

**NOTE:**

1. A wide scattering of points shows **low correlation** whereas narrow scattering shows **high correlation**. Diagram 6 shows no correlation.
2. Diagram 7 shows perfect correlation. This suggests that a formula or algebraic equation connects the two variables.
3. For a line of best fit to have much meaning the correlation must be quite high. See diagram 5.

**Exercise 1**

1. For questions a to c, draw scatter diagrams to a suitable scale, and describe each type of correlation shown.

- a) The marks for 10 pupils studying both mathematics and physics were as shown:

Maths	54	22	65	68	69	40	46	30	50	79
Physics	63	36	69	70	71	49	54	40	53	82

- b) Ten samples of home-made jam were judged separately by Mr Brown and Mr Smith and were award points as follows.

Mr Smith	30	22	25	17	17	39	33	38	27	33
Mr Brown	25	20	21	15	16	35	30	32	23	22

- c) The table shows: The number of cornets sold from an ice-cream van on each of seven days, and, the temperature at mid-day in °C.

No of cornets sold	40	70	150	50	100	170	200
Temperature	16	19	22	18	20	24	26

2. For the question below, draw a scatter diagram to a suitable scale, put in the line of best fit and use your diagram to answer the questions.

- a) The marks for ten pupils studying science were:

Physics	35	24	37	41	39	30	32	27	31	41
Chemistry	30	17	35	39	38	24	27	24	29	40

If a pupil missed the Chemistry exam, but obtained 36 marks in the Physics exam, what would be the expected mark in Chemistry?

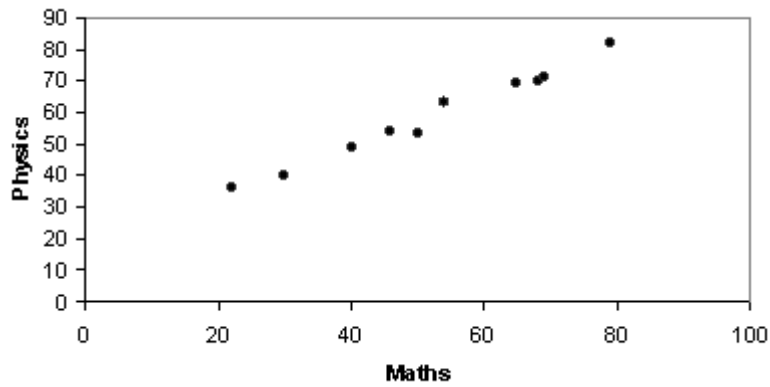
- b) Ten flower arrangers were assessed by two judges X and Y and awarded marks as follows:

Judge X	18	7	12	3	19	5	1	20	14	7
Judge Y	15	6	11	4	15	6	2	17	13	7

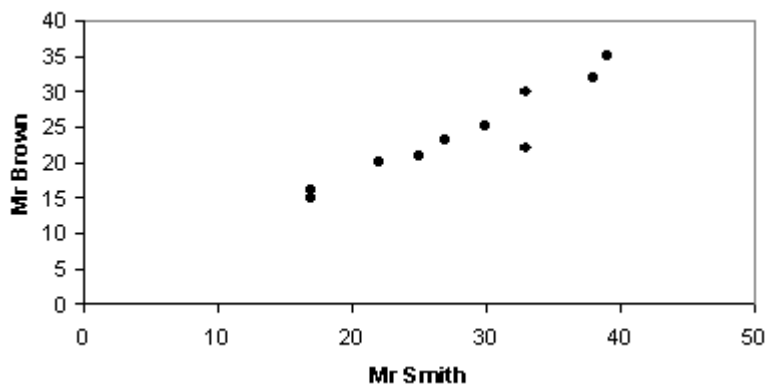
Plot these on a scatter diagram.

**ANSWERS**

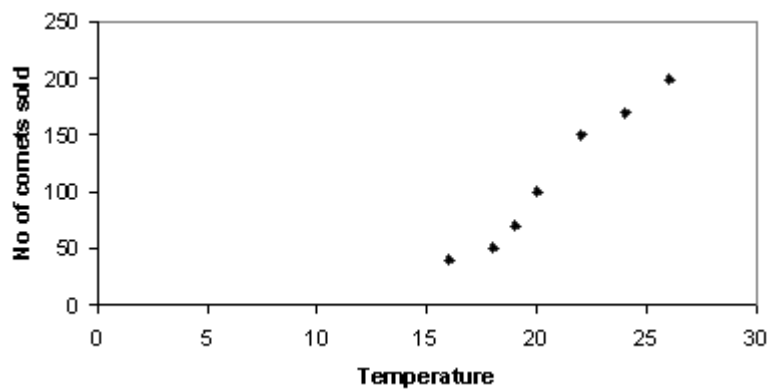
1. a) High positive correlation



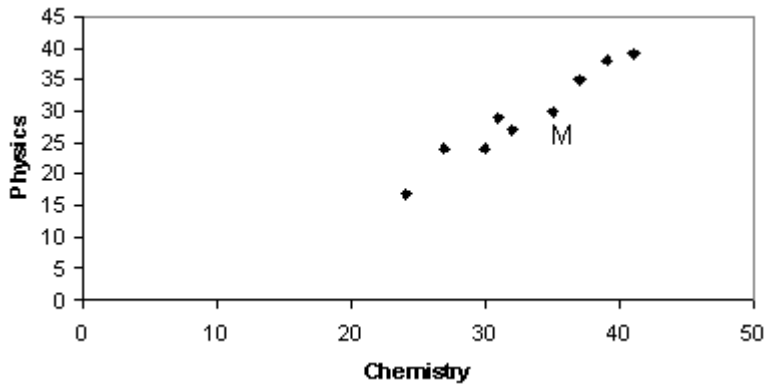
b) Positive correlation



c) Positive correlation



2. a)

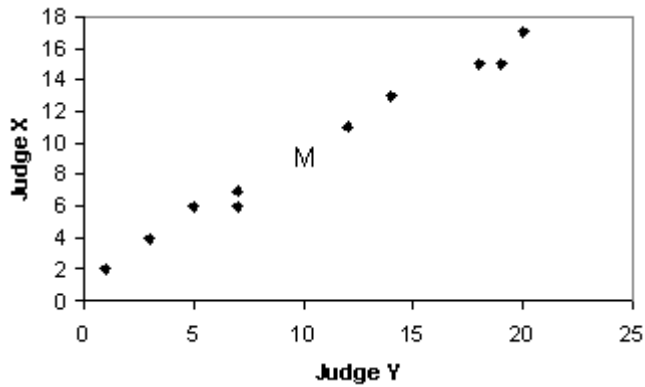


$$\text{Mean of physics marks} = \frac{337}{10} = 33.7$$

$$\text{Mean of chemistry marks} = \frac{303}{10} = 30.3$$

Pupil gaining 36 marks in physics expects 33.6 (34) in chemistry

b)



$$\text{Mean for Judge X} = \frac{106}{10} = 10.6$$

$$\text{Mean for Judge Y} = \frac{96}{10} = 9.6$$