# Scatter Diagrams.

## Please note every effort has been made to make this pack accessible but due to the nature of some of the content in this pack, the graphs and charts may not be accessible by some screen readers. If you have any queries or need any help with this pack, please email learningdevelopment@northampton.ac.uk

## 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 130cm shoe size 2

 B 135cm size 4

 C 135cm size 3

 D 155cm size 6

 E 150cm size 5

 F 160cm size 7

 G 120cm size 1

 H 170cm size 9

 I 140cm size 4

 J 125cm size 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.

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.

Scatter diagram 5 shows negative correlation or high correlation:

****

Scatter diagram 6 shows no correlation:

****

Scatter diagram 7 shows perfect positive correlation:

****

Scatter diagram 8 shows 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:

 1st Pupil Maths 54 Physics 63

 2nd 22 36

 3rd 65 69

 4th 68 70

 5th 69 71

 6th 40 49

 7th 46 54

 8th 30 40

 9th 50 53

 10th 79 82

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

 Sample Number. Mr Smith’s points. Mr Brown’s points

 No. 1 30 25

 No. 2 22 20

 No. 3 25 21

 No. 4 17 15

 No. 5 17 16

 No. 6 39 35

 No. 7 33 30

 No. 8 38 32

 No. 9 27 23

 No.10 33 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 degrees Celcius (ºC).

 Day 1 40 cornets at Temperature $16$

 Day 2 70 at 19

 Day 3 150 at 22

 Day 4 50 at 18

 Day 5 100 at 20

 Day 6 170 at 24

 Day 7 200 at 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:

 1st Pupil Physics 35 Chemistry 30

 2nd 24 17

 3rd 37 35

 4th 41 39

 5th 39 38

 6th 30 24

 7th 32 27

 8th 27 24

 9th 31 29

 10th 41 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:

 1st Flower arrangerJudge X 18 points Judge Y 15 points

 2nd 7, 6

 3rd 12, 11

 4th 3, 4

 5th 19, 15

 6th 5, 6

 7th 1, 2

 8th 20, 17

 9th 14, 13

 10th 7, 7

Plot these on a scatter diagram.

**ANSWERS.**

1. a) High positive correlation shown in the scatter diagram below:

 

b) Positive correlation shown in the scatter diagram below:

 

c) Positive correlation shown in the scatter diagram below:

 

2. a)

 

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

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

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

b)

 

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

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

This concludes the Statistics – Scatter Diagrams study pack.