How to interpret a scatterplot

Tuesday, 26 February 2019 5:56 PM

★ By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Know how to identify if there is a clear **pattern** in a scatter plot
- Know how to describe the **direction** of a scatter plot
- Know how to identify any **outliers** for a scatter plot
- Know how to describe the **form** of a scatter plot
- Know how to describe the **strength** of a relationship from a scatter plot
- Know what the correlation coefficient is

RECAP:

In previous lessons we have been looking at how to define is there is an association between various variables. In the last lesson we looked at how to draw a scatter plot (which is the best way to see if there is an association between two numerical variables.

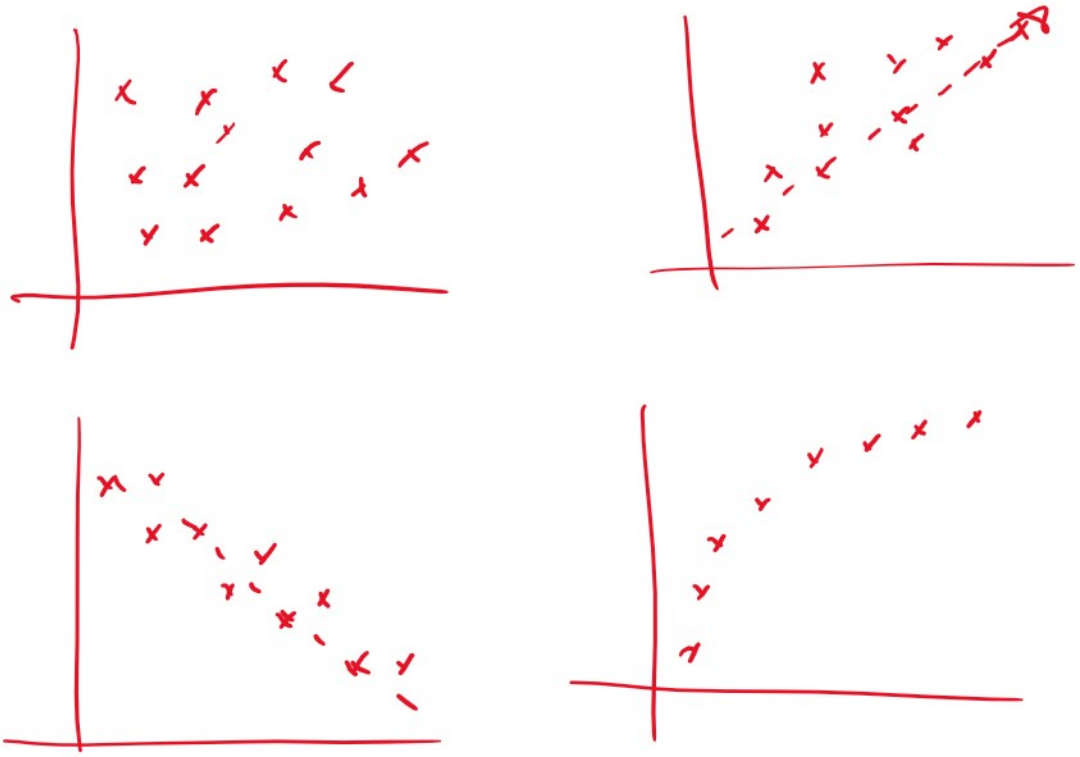
Having drawn some scatter plots, we now need to be able to describe them.

As with most things Mathematical, we need to use the correct language to be able to describe them. Reports can be written on these things!

Four main types of patterns in Scatter Plots

Shown below are the four main types of patterns which are found in scatter plots:

- No clear pattern
- Dots seemingly heading with a linear positive gradient
- Dots seemingly heading with a linear negative gradient
- Dots which seem to have a non-linear association



Making sure we describe Scatter plots with all the key terms

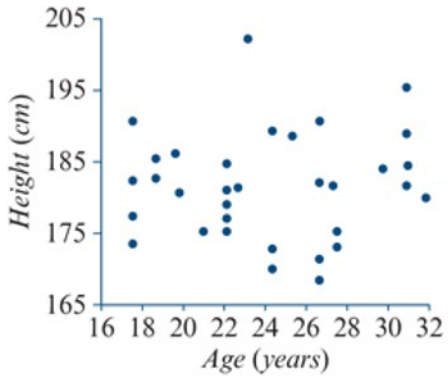
It is vital, in Further Maths, to describe scatter plots using three (or four) main terms:

- Direction
- Outliers (if any)
- Form, and
- Strength

Let's talk about Direction

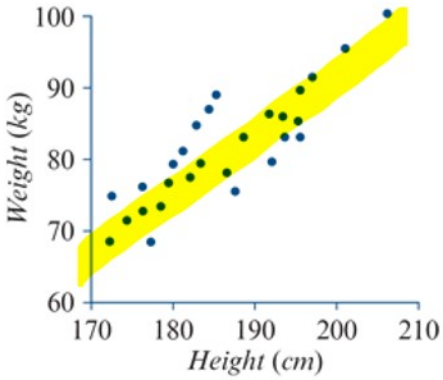
OK! Not One Direction ... But the direction the crosses (or dots) seem to be heading.

Random scatter of points: No association



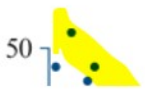
No consistent change in the value of the response variable when the values of the explanatory variable increases.

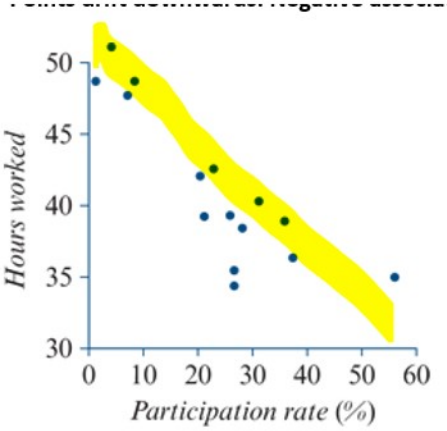
Points drift upwards: Positive association



The response variable tends to increase as the value of the explanatory variable increases

Points drift downwards: Negative association





The response variable tends to decrease as the value of the explanatory variable increases

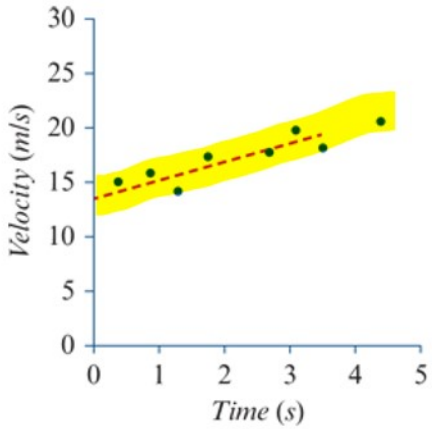
→ PR ↑ HW ↓

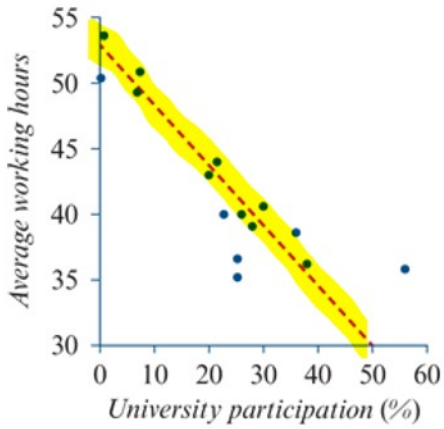
Let's talk about Form

This is basically describing the drift as either:

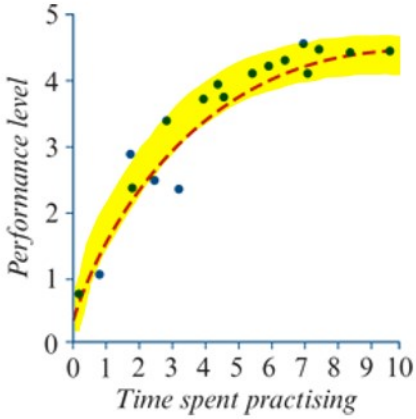
- Linear, or
- Non-linear

So, looking at three examples:





The above two graphs suggest a linear association



This graph suggests a non-linear association (as the points seem to curve).

Strength of a linear relationship

It's important to now get very used to a new term: **correlation coefficient**.
This is a measure of how close the points are on a scatter plot to a "line of best fit".

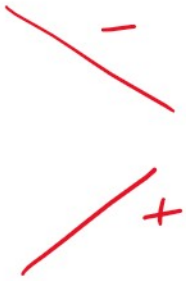
The correlation coefficient is a measure of how linear a relationship is.

We use the letter r to stand for the correlation coefficient.

It takes a value between -1 and 1.

Pearson's

Negative values are used for negative linear associations.
Positive values are used for positive linear associations.
A zero value means there is no association.



There is a helpful table of results which can help us turn a value of r into a english phrase.
This is helpful when writing reports.

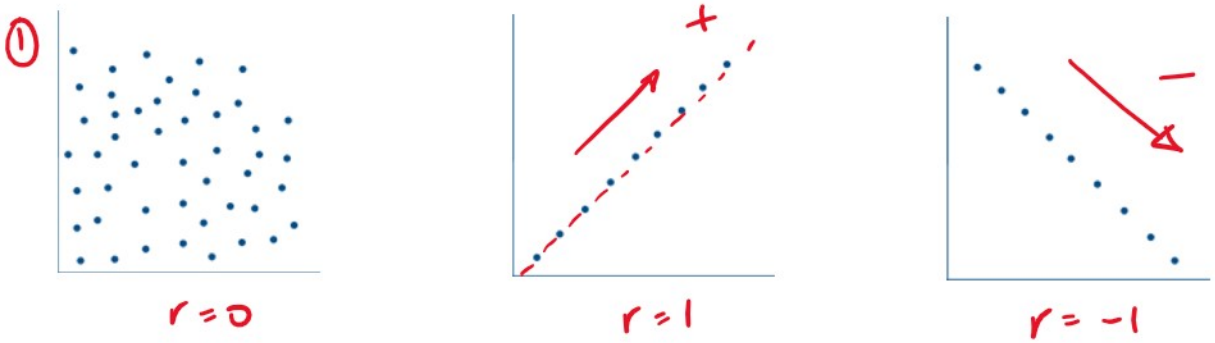
Strong positive association: r between 0.75 and 0.99
Moderate positive association: r between 0.5 and 0.74
Weak positive association: r between 0.25 and 0.49
No association: r between -0.24 and $+0.24$
Weak negative association: r between -0.25 and -0.49
Moderate negative association: r between -0.5 and -0.74
Strong negative association: r between -0.75 and -0.99

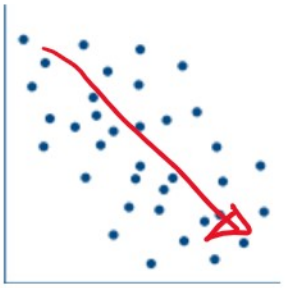
This table will be really important when we start calculating the value of r in questions.

Notice the use of the words:

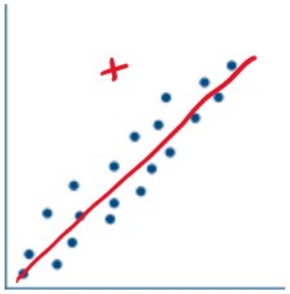
- Weak
- Moderate
- Strong

Examples of graphs of linear relationships and their values of r.





$r = -0.3$



$r = 0.9$

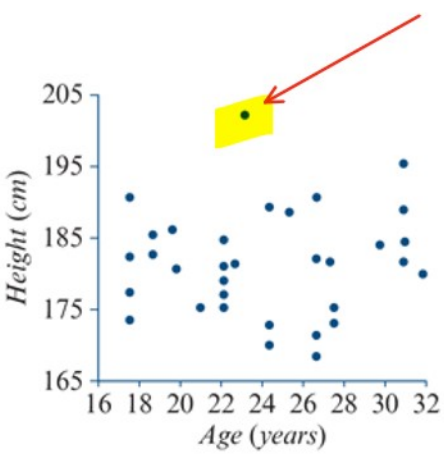
Important information:

When using the correlation coefficient we are assuming the following:

- Variables are numeric
- Association is linear
- There are no outliers

Hold on ... What's an outlier?

If we look back at one of the first graphs we can see a point which seems quite separate from the others!



VCAA Exam Question on this concept
2016 Paper 2

Question 3 (8 marks)

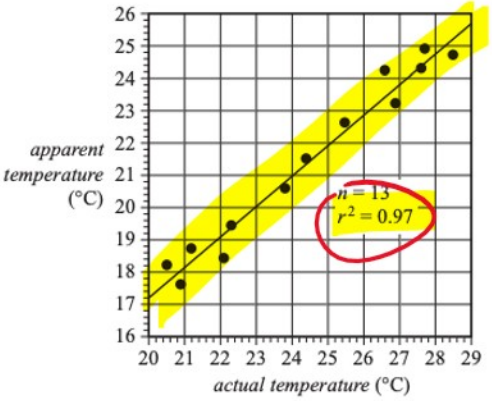
The data in the table below shows a sample of actual temperatures and apparent temperatures recorded at the weather station. A scatterplot of the data is also shown.

Question 3 (8 marks)

The data in the table below shows a sample of actual temperatures and apparent temperatures recorded at the weather station. A scatterplot of the data is also shown.

The data will be used to investigate the association between the variables *apparent temperature* and *actual temperature*.

Apparent temperature (°C)	*Actual temperature* (°C)
24.7	28.5
24.3	27.6
24.9	27.7
23.2	26.9
24.2	26.6
22.6	25.5
21.5	24.4
20.6	23.8
19.4	22.3
18.4	22.1
17.6	20.9
18.7	21.2
18.2	20.5



r = 0.98
strong positive linear

a. Use the scatterplot to describe the association between *apparent temperature* and *actual temperature* in terms of strength, direction and form. 1 mark