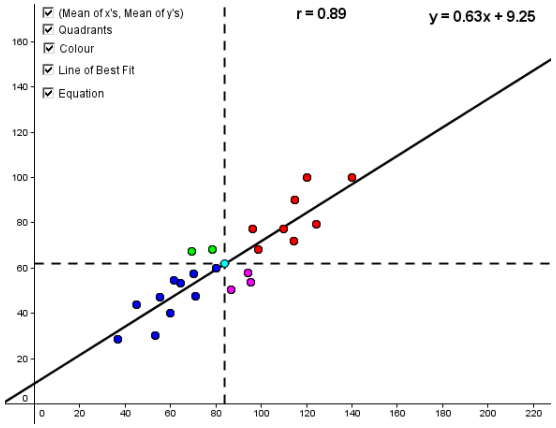
Tasks Related to "Correlation and Line of Best Fit 2"

Use in connection with the interactive file, 'Correlation and Line of Best Fit 2', on the Student's CD.

Calculating the point (Mean of x s, Mean of y s) and using it to split the plane into 4 quadrants is <u>not</u> specifically mentioned in the syllabus. However, it is a helpful way of learning about relationships in the data.

The purpose of this task sheet is to explore the relationship between the pattern of the points in a scatter plot, the correlation coefficient, line of best fit, outliers, the point (Mean of x s, Mean of y s) and the numbers of points in each quadrant.



Task 1

Leave the 24 points as they are. Click on **"(Mean of x s, Mean of y s)"**. This shows the point that is the centre of the data. We will call this the point A

Click on **"Quadrants"**.

<u>Complete the sentence:</u> The bottom left quadrant has all the points that have below average x-values and _____ average y-values.

<u>Complete the sentence:</u> The top right quadrant has all the points that have above average x-values and _____ average y-values.

Click on **"Colour"**.

Count up all the points in each of the quadrants. Does the amount of points in each quadrant hint at a relationship in the data?

By looking at your answers to the previous questions and the diagram is there a **<u>linear</u>** relationship in the data?

By looking at both the correlation coefficient and the scatter plot could you say that as the x-values of the points increase the y-values of the points tend to increase?

Click on **"Line of Best Fit"** and **"Equation"**.

Does the line of best fit pass through all the points in the scatter plot?

Does the line of best fit have to go through any of the points in the scatter plot?

What can you say about the point A and the line of best fit?

Your friend George asks "Does the line of best fit have to have half the points on either side of it?". Move the points and see if you can answer George's question.

Drag some of the points around the screen so that the linear relationship is maintained. Which 2 quadrants does the line of best fit pass through when the correlation coefficient (r) is close to 1?

Drag some of the points around the screen so that the points are still in a linear relationship. Which 2 quadrants does the line of best fit pass through when the correlation coefficient (r) is close to –1?

Task 2

Reset the scatter plot using the icon at the top right-hand corner of the screen  .

Move the points so that there are an equal number of points in all four quadrants.

(a) Arrange the points so that they are bunched together into a "cloud" or "swarm". Is the correlation coefficient close to 1, –1 or 0?

(b) Arrange the points so that they resemble the outline of a circle. Is the correlation coefficient close to 1, –1 or 0?

(c) Arrange it so that the points in two of the quadrants are stretched out into a line and the points in the other two quadrants are close to the point A is the correlation coefficient close to 1, –1 or 0?

Task 3

Reset the scatter plot using the icon at the top right-hand corner of the screen  .

Put 10 points in the top right quadrant. Put 10 points in the bottom left quadrant.
Put 2 in the top left quadrant. Put 2 points in the bottom right quadrant.

While keeping the number of points in each quadrant as outlined above can you adjust the points so that there is a correlation close to -1 or 1?

While keeping the number of points in each quadrant as outlined above can you adjust the points so that there is a correlation close to 0?

Having a large number of points in the bottom left and top right quadrants doesn't always indicate association between the variables. There must be a linear pattern for the correlation coefficient (r) to be close to -1 or 1.

Task 4

Reset the scatter plot using the icon at the top right-hand corner of the screen  .

Move the points so that the pattern looks like a quadratic.

Is there a pattern to the points?

Is the pattern linear?

There may be a strong association between the variables, but since the relationship is not linear it wouldn't be useful to summarise the strength of the relationship with the correlation coefficient (r) or to draw a line of best fit.

Task 5

Reset the scatter plot using the icon at the top right-hand corner of the screen  .

Group all the points into a tight bunch in one of the corners of the screen.

Adjust the points until you have a correlation coefficient close to 0.

Drag one point very far away from this bunch e.g. to towards the opposite corner of the screen.

Watch the correlation coefficient changing.

23 of the points are in a bunch and there is 1 point far away from the rest.

Is there a linear relationship between the points?

What conclusion can you draw about the effect of the outlier on the level of correlation?

The correlation coefficient indicates a strong linear relationship but by looking at the graph you see that the relationship is not linear (without the outlier the correlation coefficient is near 0).

It is important to analyse the data both numerically (correlation coefficient) and graphically (scatter plot).

A single outlier can bring the value of r close to –1 or 1

Task 6

Set the points up so that they are (almost) in a line (that isn't horizontal or vertical).

Take note of the correlation coefficient (r).

Drag one point very far away from this line of dots.

How does the correlation coefficient (r) change?

A single outlier can bring the correlation coefficient close to zero.