

Using the least squares line to model a linear association

Year 11 General Maths
Units 1 and 2

www.maffsguru.com

Learning Objectives

By the end of the lesson I hope that you understand and can apply the following to a range of questions from the Unit 1 and 2 General Mathematics course.

- Understand what is meant be the term least squares regression line
- Understand what a residual is
- Know what the equation of a least squares regression line is
- Know how to find the least-squares regression line using the CAS

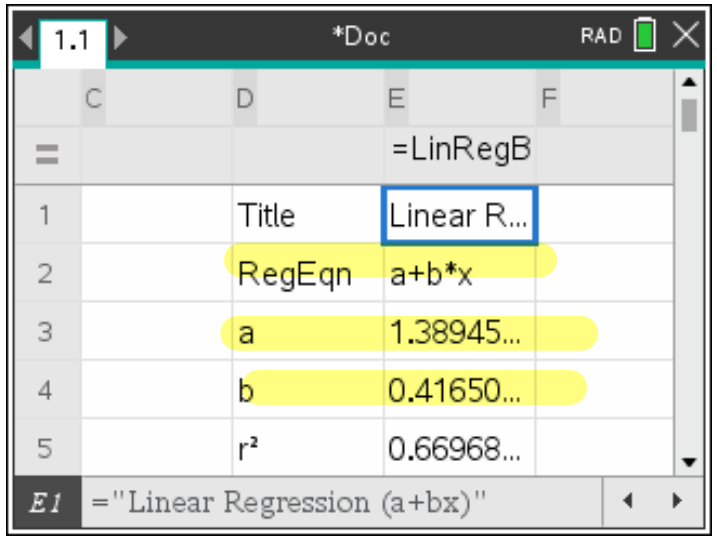


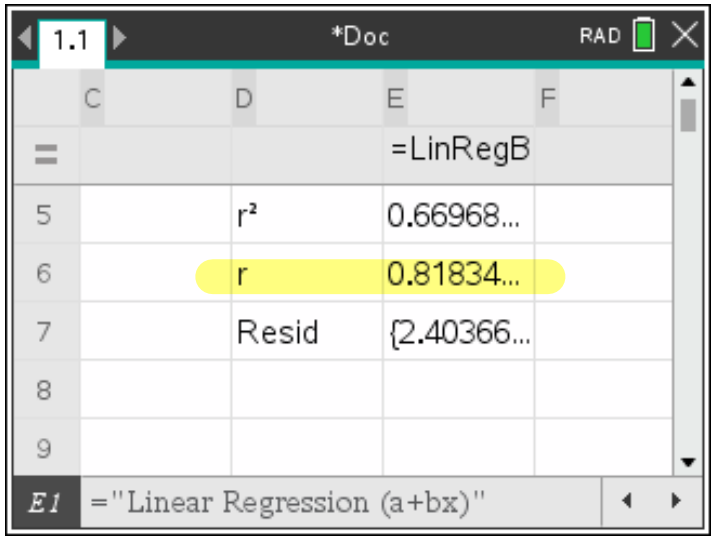
Recap of past learning

Firstly, the good news … you already know how to find the information which will lead to the equation of a least squares regression line. You covered it in the previous lesson. The last two screens I came up with (shown below) have all the information you need.

It is important that you have an understanding of how the values are made!

There is a **pretty funky** formula which I have seen appear on lots of Year 12 exam papers and Year 11 end of year exams.







What is a least squares regression line?

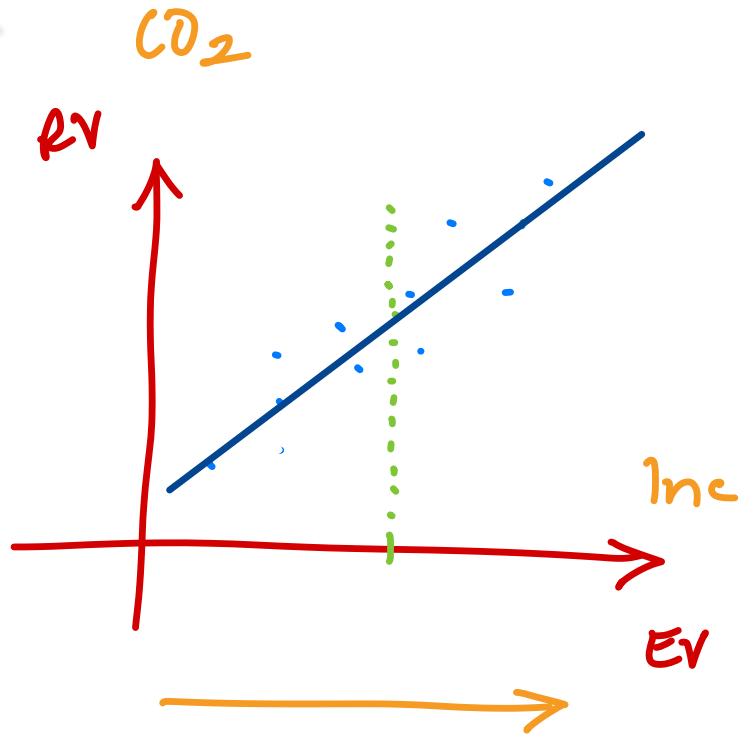
Basically, it's a line of best fit which passes through the points of a scatter plot.

It is there to help us predict values which we might not have data for.

Because the least squares line is "a work of mathematical fiction" then we are going to have some errors when we predict values outside of those we have been given.

So, if we go back to the previous example, what would the CO_2 output be for an income of $20,000?

Income ($000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 (tonnes)	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9

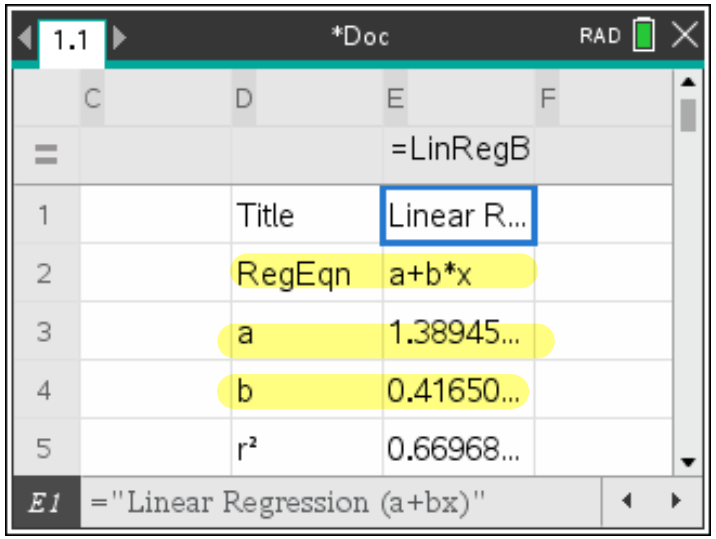


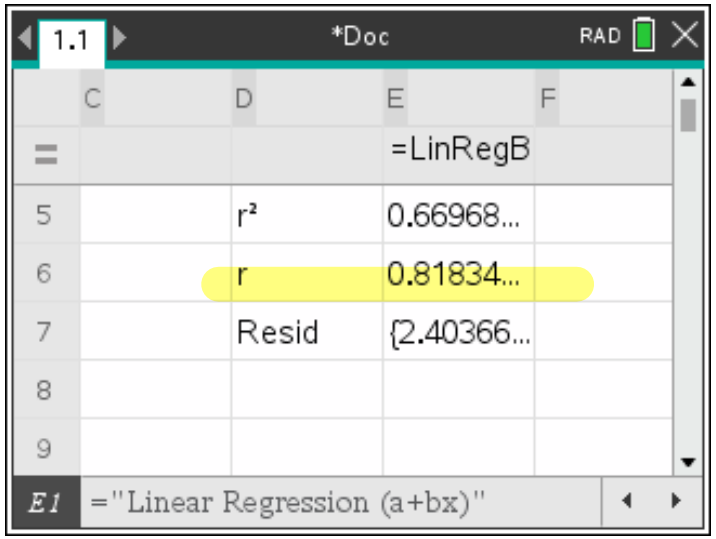


Finding the equation of the least squares line

$y = mx + c$

Let's use the same example we used in the previous video!
Remember, that the two final screens showed the following:





Do you notice the values of 'a' and 'b' on the first screen?



Moving the letters around

In General Maths we don't use the traditional equation of a straight line which you have used since Year 9:

Namely, $y = mx + c$

We now use: $y = a + bx$

But what are the 'a' and 'b'?

$y = mx + c$

$y = a + bx$

$a =$ int

$b =$ grad

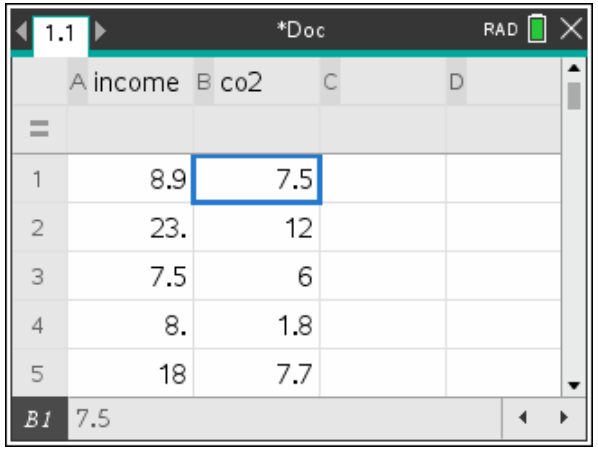


Each value in a table is a coordinate!

It is important to know that the point in the table are nothing more than coordinates on the graph.

You can get a regression line from as few as two coordinates (which becomes really important later in this lesson).

Income ($000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 (tonnes)	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9





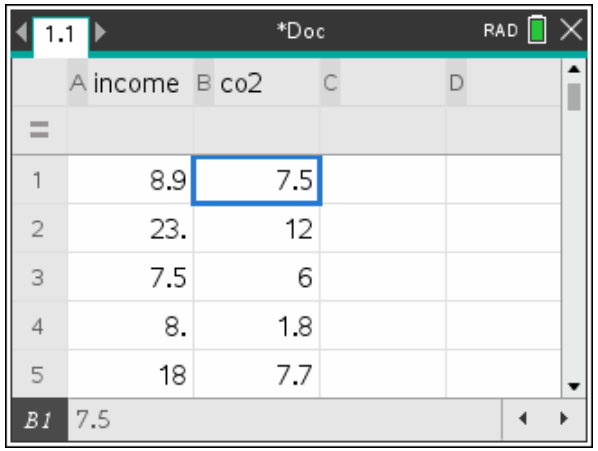
Using the graph drawing capabilities

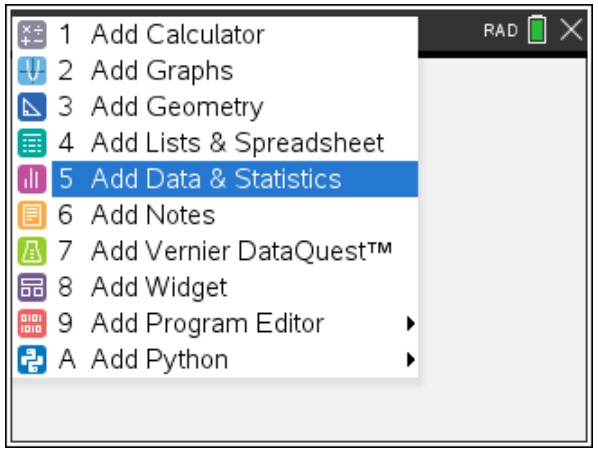
We can also use the graph drawing functions of the CAS to give me the same information.

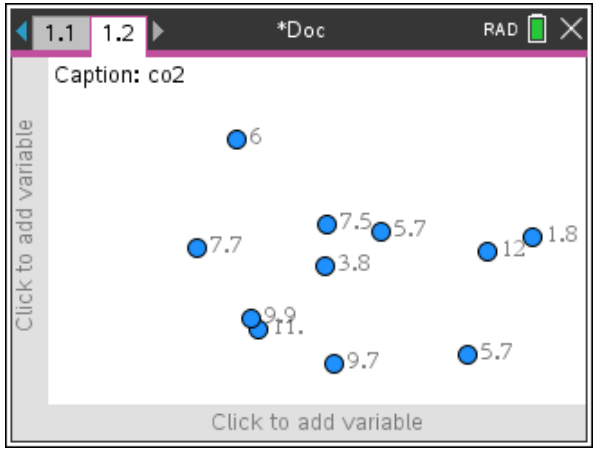
The following data shows the per capita income (in $000) and the carbon dioxide emissions (in tonnes) of 11 countries.

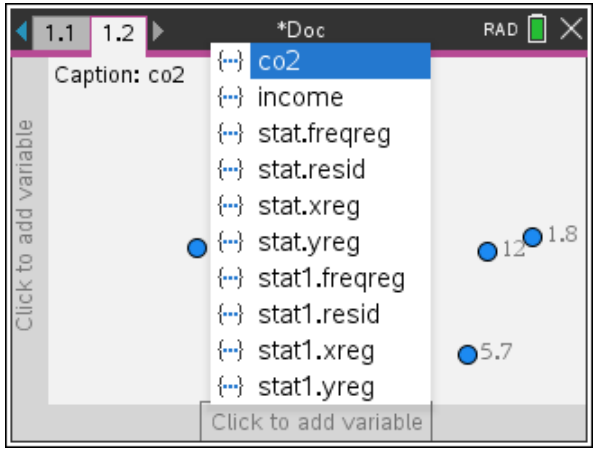
Determine the value of Pearson's correlation coefficient, r, for these data.

Income ($000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 *(tonnes)*	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9











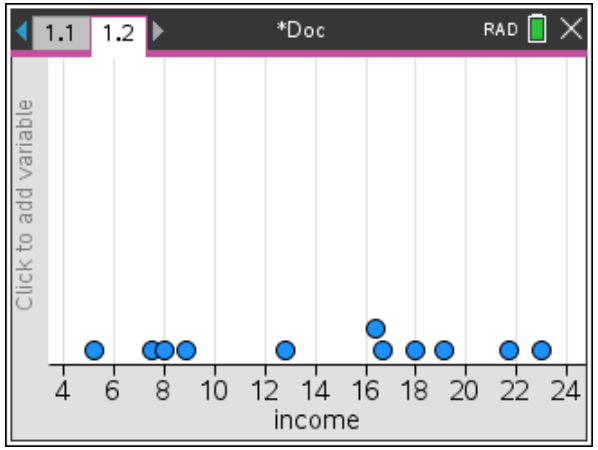
Using the graph drawing capabilities

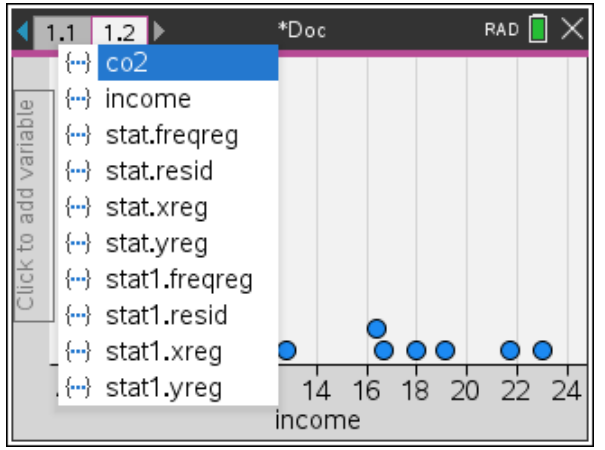
We can also use the graph drawing functions of the CAS to give me the same information.

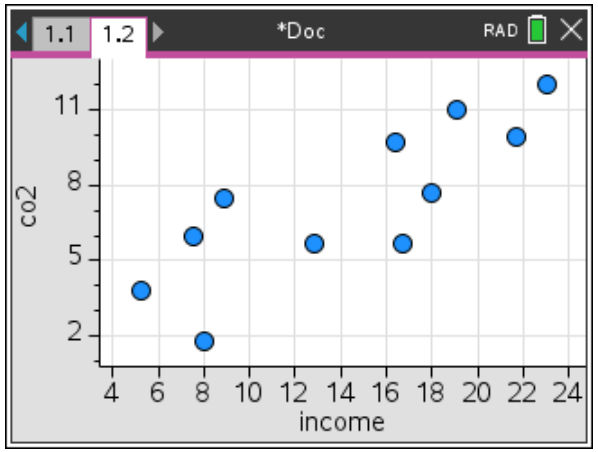
The following data shows the per capita income (in $000) and the carbon dioxide emissions (in tonnes) of 11 countries.

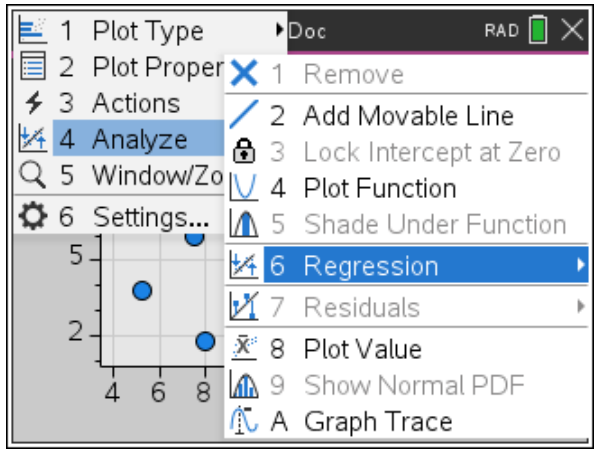
Determine the value of Pearson's correlation coefficient, r, for these data.

Income ($000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 *(tonnes)*	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9











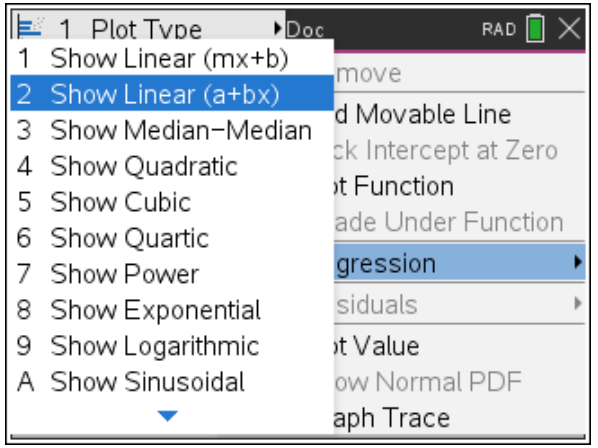
Using the graph drawing capabilities

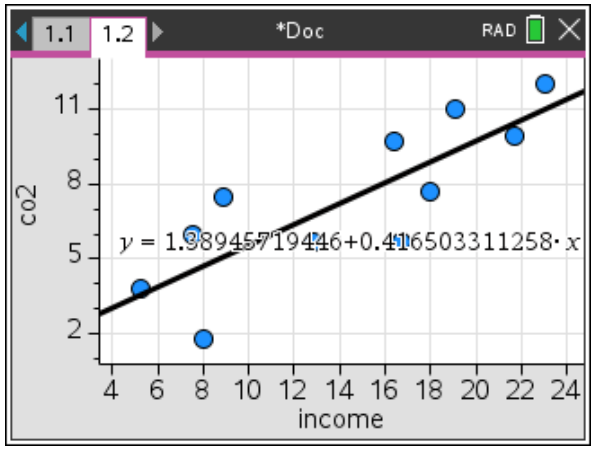
We can also use the graph drawing functions of the CAS to give me the same information.

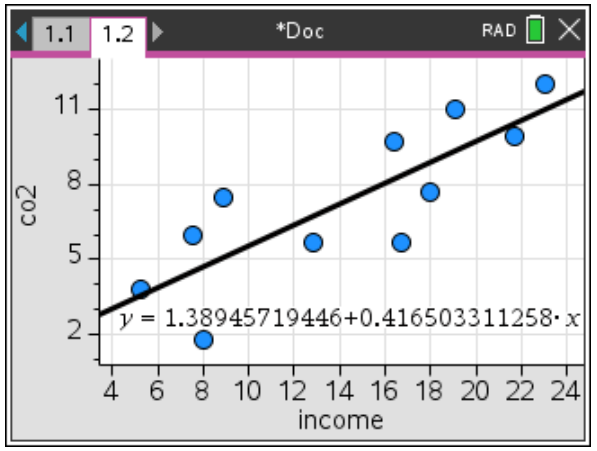
The following data shows the per capita income (in $000) and the carbon dioxide emissions (in tonnes) of 11 countries.

Determine the value of Pearson's correlation coefficient, r, for these data.

Income ($000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 (tonnes)	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9







BEWARE THE TRICK OF THE EQUATION …

Your CAS is a dumb piece of machinery.

It gives you the line with 'y' and 'x' in it. BUT you must change those for the actual variables you used!! In my case, Income and CO_2



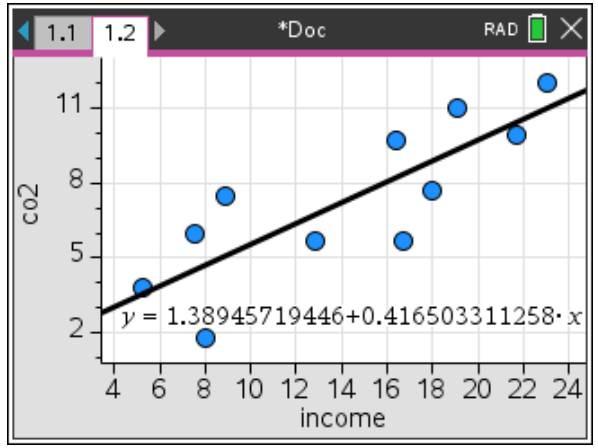
Finding the ‘r’ value

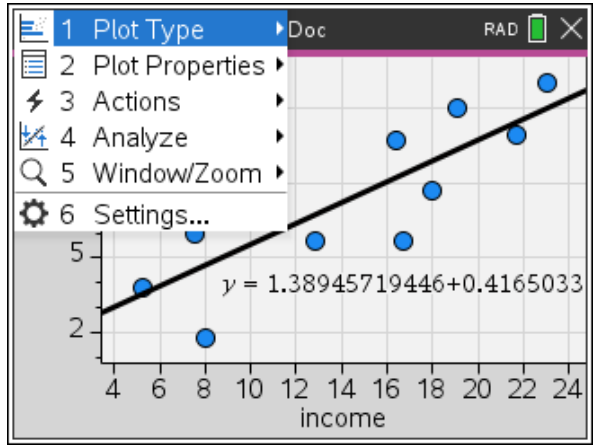
There are times we will also be asked to find the value of ‘r’.

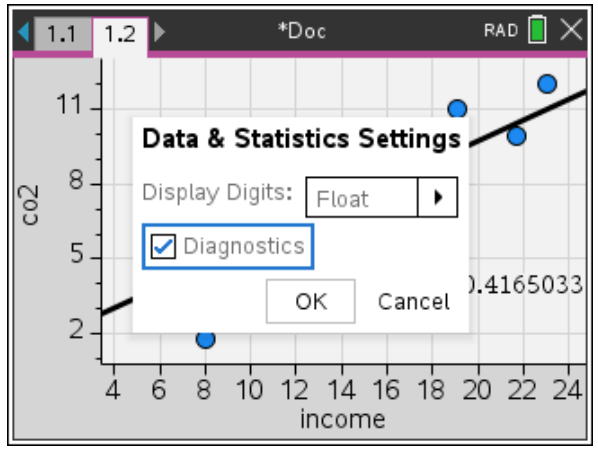
This is relatively easy to do.

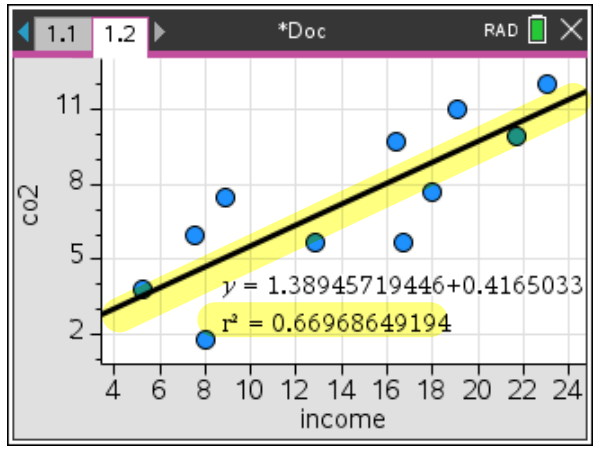
We can get the value of r^2 which is going to be helpful in another video … and the square root it to find the value of ‘r’

Income ($000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 (tonnes)	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9





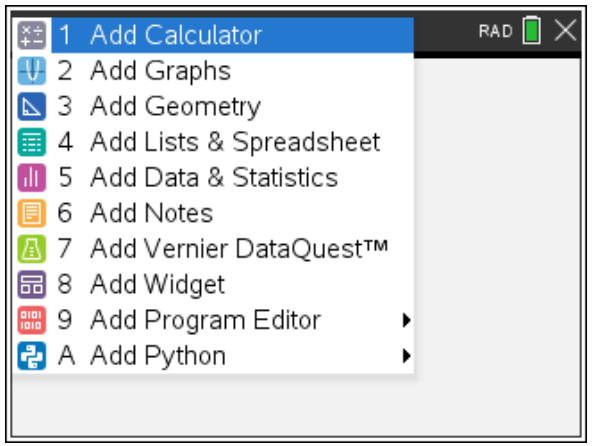


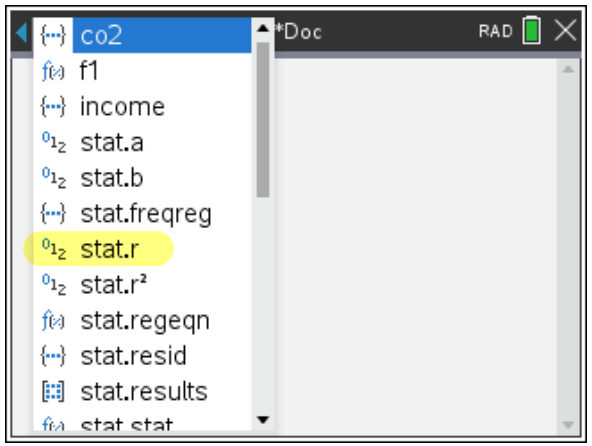


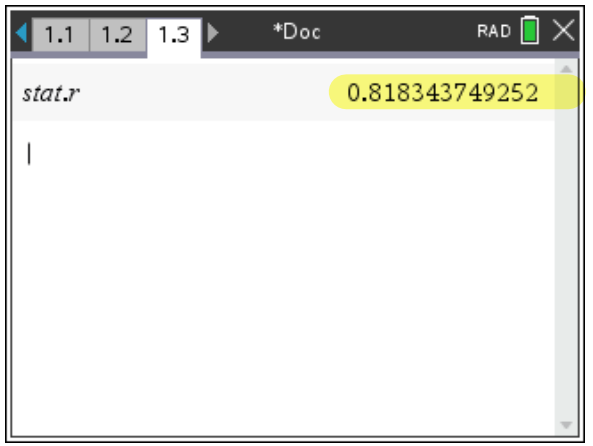


Finding the ‘r’ value

Or … we can use a CAS function :







$r^2 = 4$

$r = \sqrt{4}$

$= \pm 2$

VAR

$r^2 = \sqrt{r}$

$2 \times 2 = 4$

$-2 \times -2 = 4$





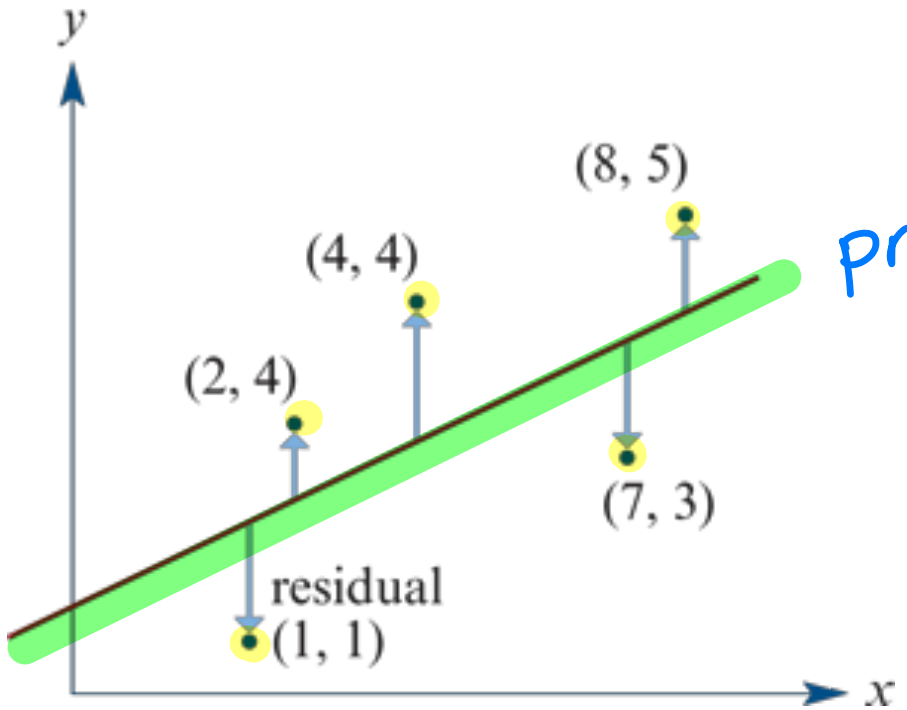


Residual values

We can get the CAS to take the pieces of information given (data) and guess where a line of best fit might be.

One way it does this is using a method called finding the least squares.

Before we do this … let's look at something called a residual value. This helps us understand how the CAS does the magic it does.

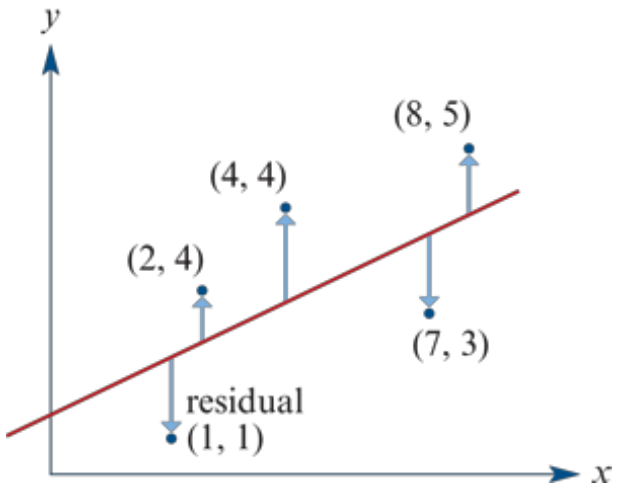


predict RV from EV



Making the values of the residual as small as possible

Using the idea below ...





You mentioned an equation...

Basically, the CAS uses the following to allow it to find the equation of the least squares regression line:

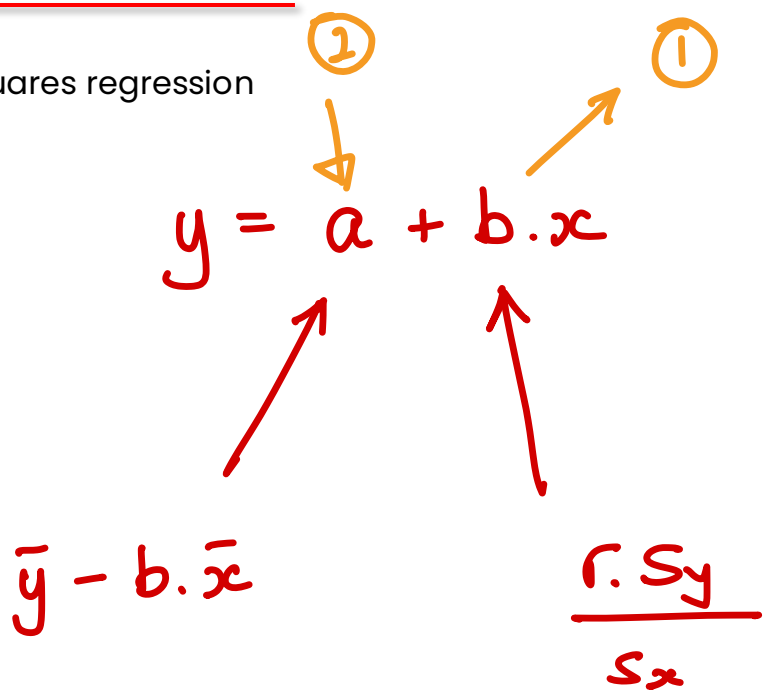
Knowing the equation of a straight line can be written as $y = a + bx$:

The slope is given by:

$$b = \frac{r\, s_y}{s_x}$$

The intercept is given by:

$$a = \bar{y} - b\bar{x}$$





Example of using the formula

Find the equation of the least squares regression line, $y = a + bx$, when:

$r = 0.600$ $s_x = 9.80$ $s_y = 5.40$ $\bar{x} = 165.0$ $\bar{y} = 62.3$

Give the values of the intercept and slope, rounded to three significant figures.

$y = a + b \cdot x$

$b = \frac{r \cdot s_y}{s_x}$

$= \frac{0.600 \times 5.40}{9.80}$

$= 0.331$

$= \bar{y} - b \cdot \bar{x}$

$= 62.3 - 0.331 \times 165$

$= 7.69$

$\therefore$ int $= 7.69$

slope $= 0.331$



Beware Will Robinson …

Please note that I have seen the formula's on the previous page used over and over and over again in the General Maths exam. So, you must know how to use the equations given.



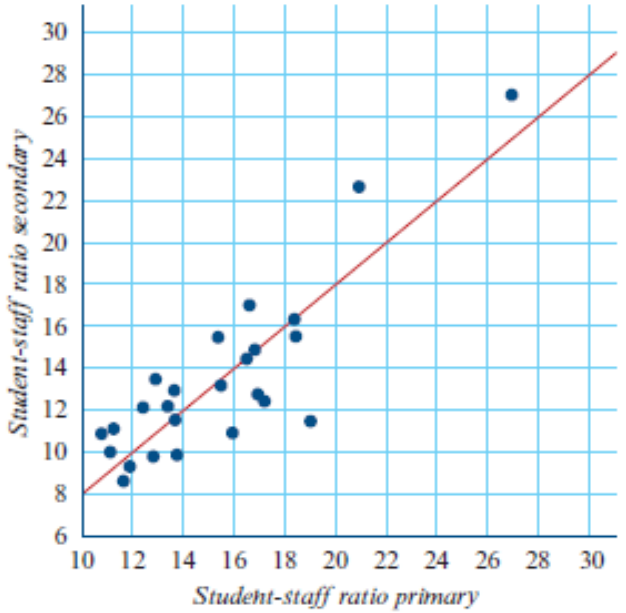
Finding the equation of a least squares line from a graph

This is a skill which most Year 12 students simply cannot do. So, learning this now and learning it well will see you have a competitive advantage next year!

To find the equation of a least squares line you simply need two points which sit on the line. Knowing this, you can then use the same process as before.

A straight line has been fitted by eye to a scatterplot of the *student-staff ratio for secondary* against the *student-staff ratio for primary* for a group of countries.

Determine the equation of this line in terms of the variables in the question.



(10,8)

(28,26)

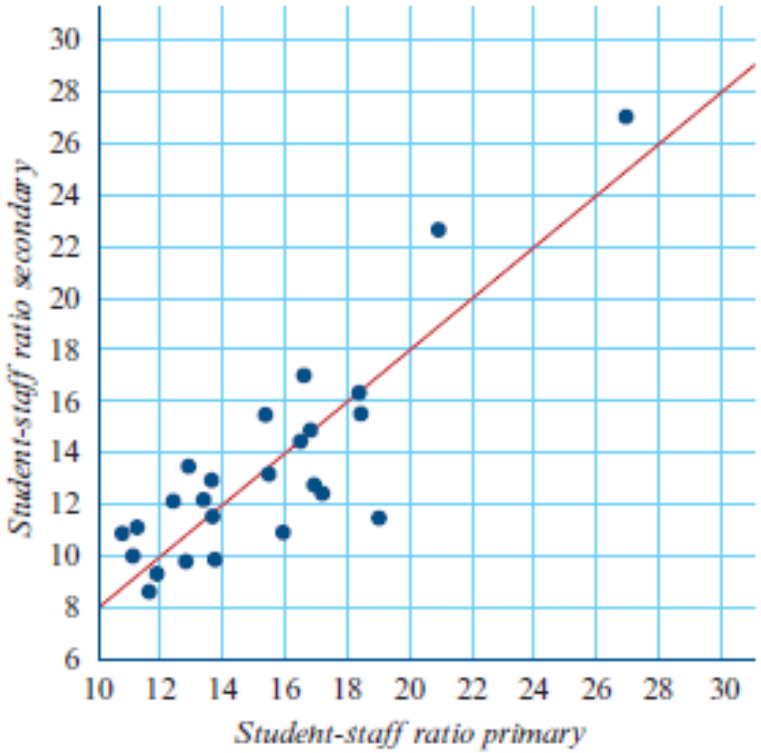
$y = -2 + x$



Finding the equation of a least squares line from a graph

A straight line has been fitted by eye to a scatterplot of the *student-staff ratio for secondary* against the *student-staff ratio for primary* for a group of countries.

Determine the equation of this line in terms of the variables in the question.



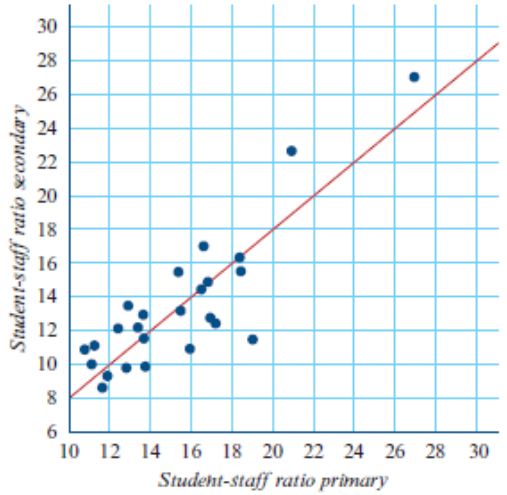
The most important thing is to choose the best two coordinates!

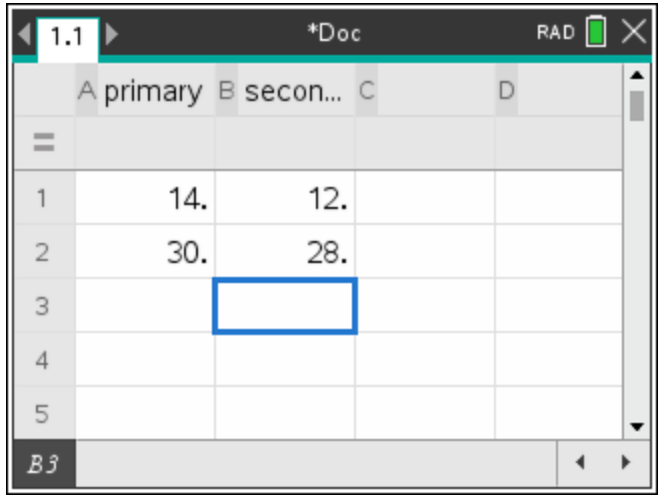


Finding the equation of a least squares line from a graph

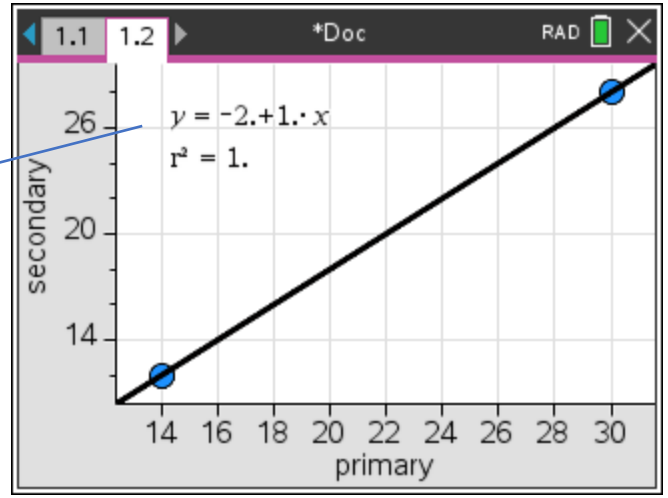
A straight line has been fitted by eye to a scatterplot of the *student-staff ratio for secondary* against the *student-staff ratio for primary* for a group of countries.

Determine the equation of this line in terms of the variables in the question.





This has been gained using the instructions used in the previous part of the lesson.





Learning Objectives: Revisited

By the end of the lesson, I hope that you understand and can apply the following to a range of questions from the Unit 1 and 2 General Mathematics course.

- Understand what is meant be the term least squares regression line
- Understand what a residual is
- Know what the equation of a least squares regression line is
- Know how to find the least-squares regression line using the CAS



Questions to complete

It is expected that, to improve your skills, you will complete the following questions:

Exercise 7D: All Questions

