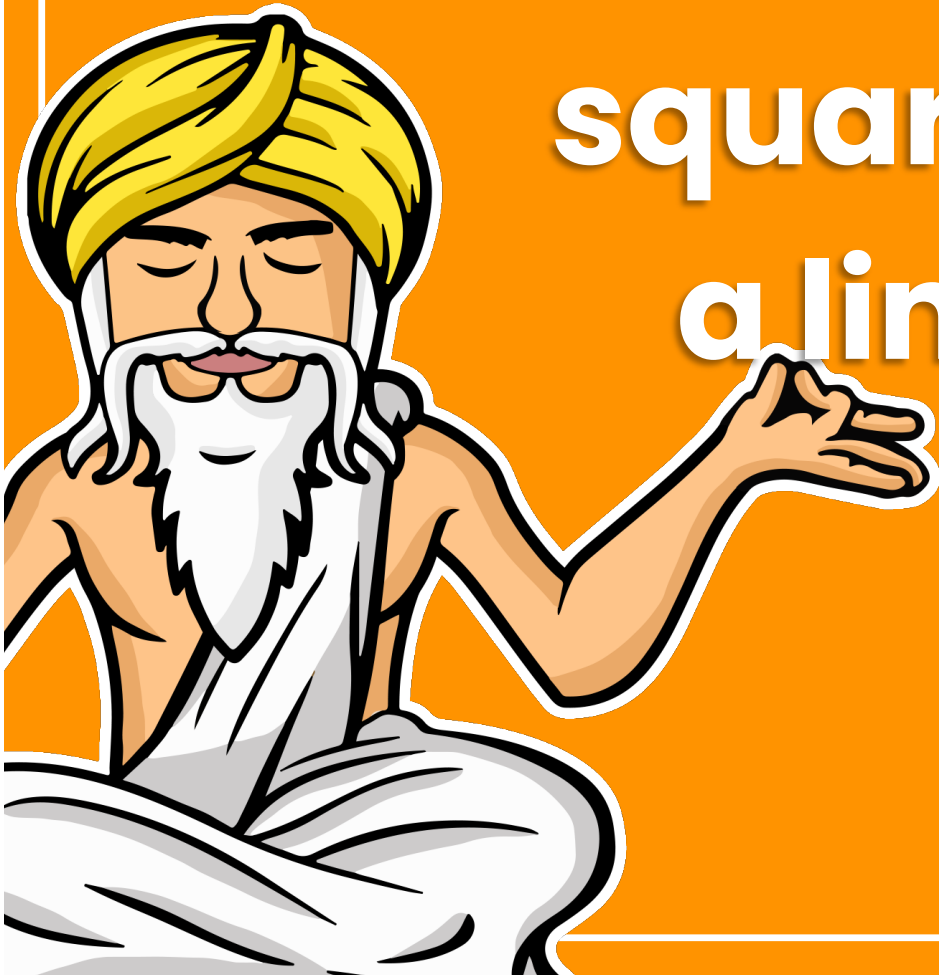


Using the least squares line to model a linear association



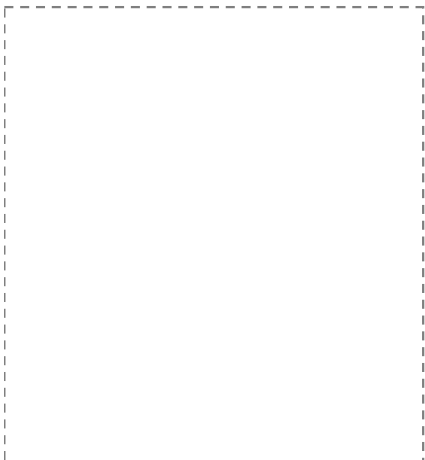
**Year 11
General Mathematics**

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 by 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

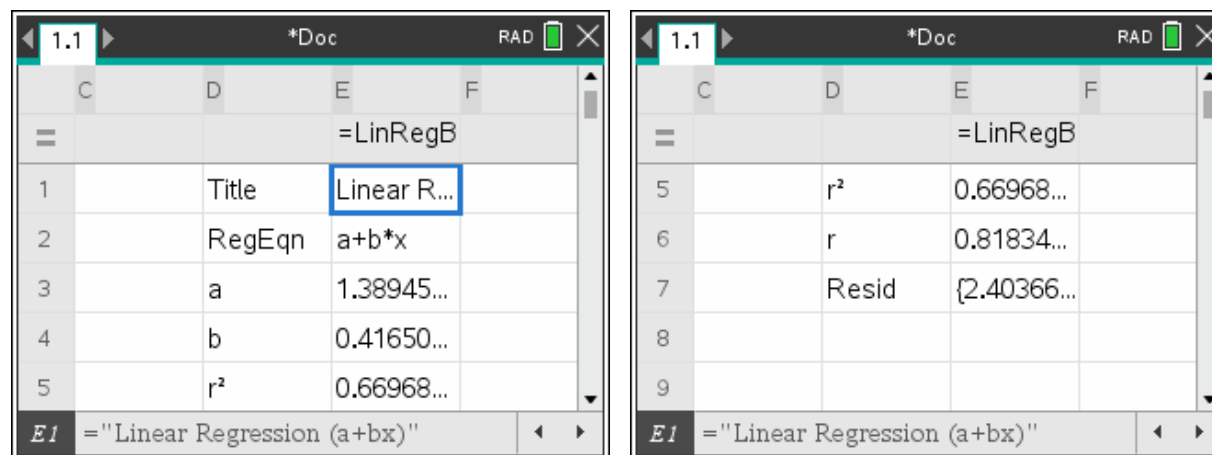


Recap of past learning

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.



	C	D	E	F
=			=LinRegB	
1		Title	Linear R...	
2		RegEqn	a+b*x	
3		a	1.38945...	
4		b	0.41650...	
5		r ²	0.66968...	
E1	="Linear Regression (a+bx)"			

	C	D	E	F
=			=LinRegB	
5		r ²	0.66968...	
6		r	0.81834...	
7		Resid	{2.40366...	
8				
9				
E1	="Linear Regression (a+bx)"			

What is a least squares regression line?

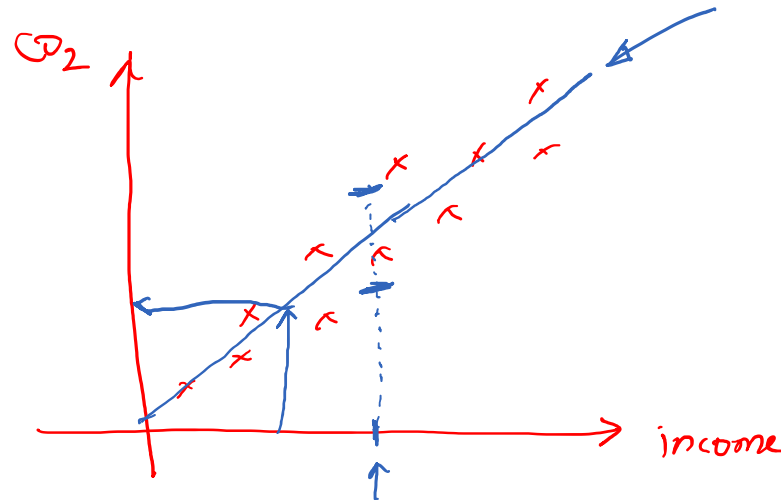
Basically, it's a line of best fit which passes through the points of a scatter plot.

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

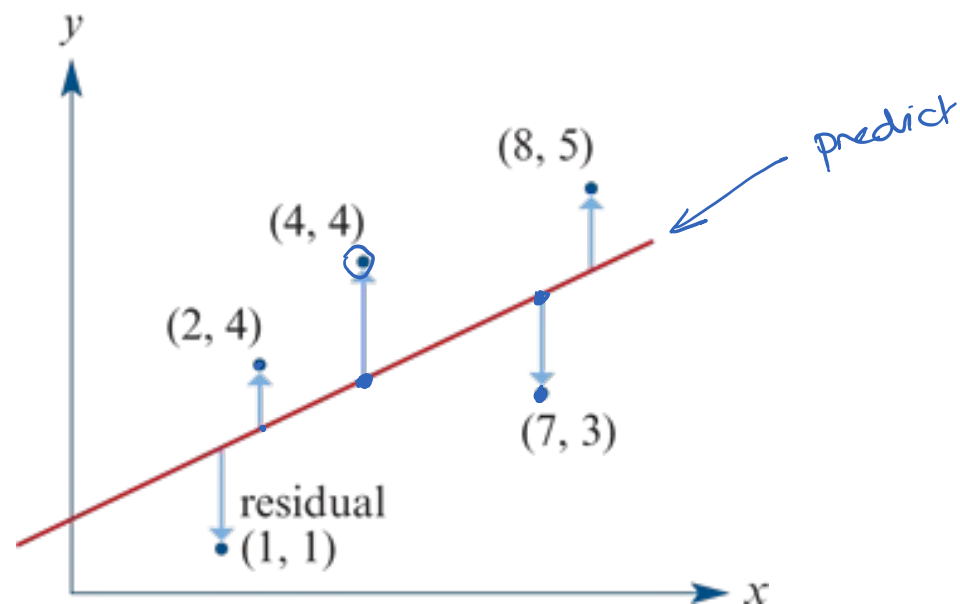


Residual values

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

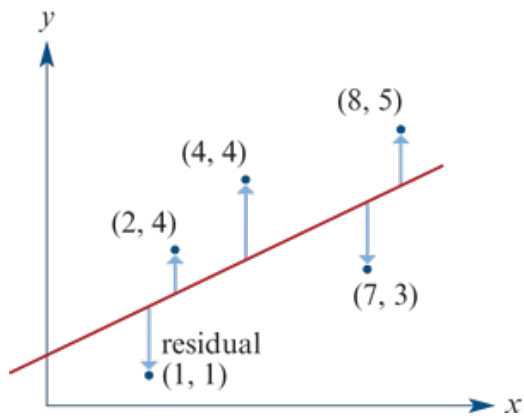
One way it does this is using a method called finding the least squares.

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.



Making the values of the residual as small as possible

Using the idea below ...



Phew! So, what's the equation then?

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 $y = a + bx$:

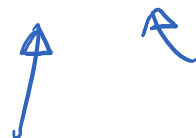
The slope is given by:

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

The intercept is given by:

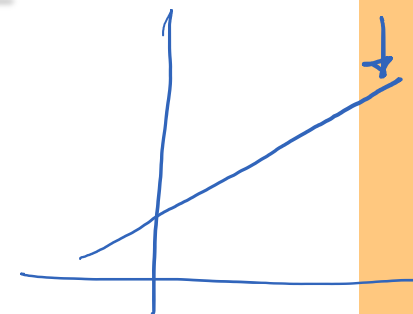
$$a = \bar{y} - b\bar{x}$$

$$y = a + bx$$



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

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



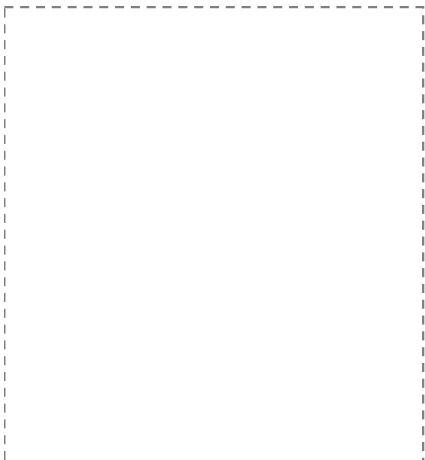
$$y = mx + c$$

$$y = a + bx$$

↑ ↗ grad
intercept

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 Further Maths exam. So you must know how to use the equations given.



An example

We can use the same example we used in the previous video!
Remember, that the two final screens showed the following:

	C	D	E	F
=			=LinRegB	
1		Title	Linear R...	
2		RegEqn	$a+b*x$	
3		a	1.38945...	
4		b	0.41650...	
5		r^2	0.66968...	
E1 = "Linear Regression (a+bx)"				

	C	D	E	F
=			=LinRegB	
5		r^2	0.66968...	
6		r	0.81834...	
7		Resid	{2.40366...	
8				
9				
E1 = "Linear Regression (a+bx)"				

$$y = 1.389 + 0.417x$$

$$\underline{\underline{CO_2 = 1.389 + 0.417 \times inc}}$$

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

These are the values of 'a' and 'b' in $y = a + bx$

So, we can use the process described in the previous video and just use the values of "a' and 'b'.

$$\begin{array}{l} CO_2 \\ \downarrow \\ inc \end{array}$$

Using the graph drawing capabilities

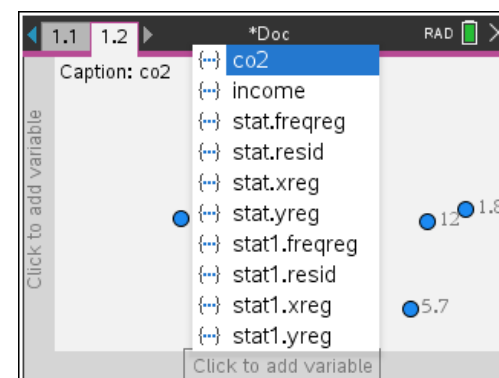
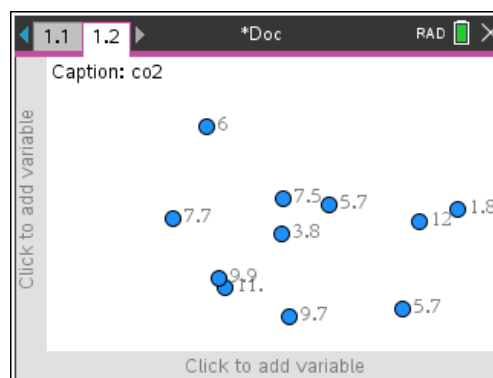
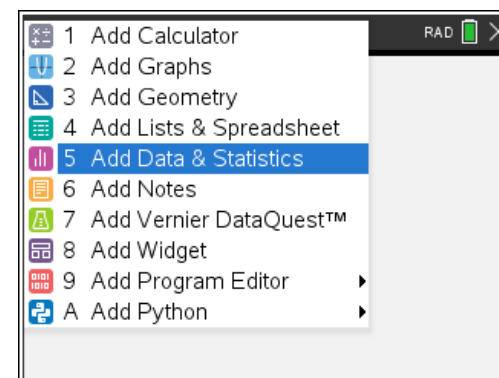
We can also use the graph drawing functions of the CAS to give me the same information.

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

	A income	B co2	C	D
1	8.9	7.5		
2	23.	12		
3	7.5	6		
4	8.	1.8		
5	18	7.7		



Examples have been extracted, with permission, from the Cambridge General Mathematics Units 1 and 2 Textbook

www.maffsguru.com

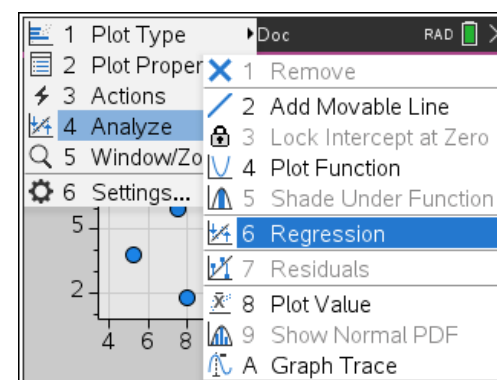
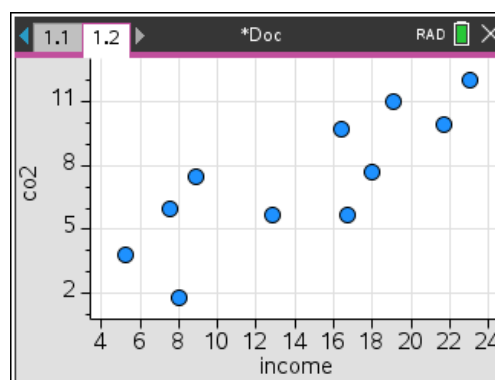
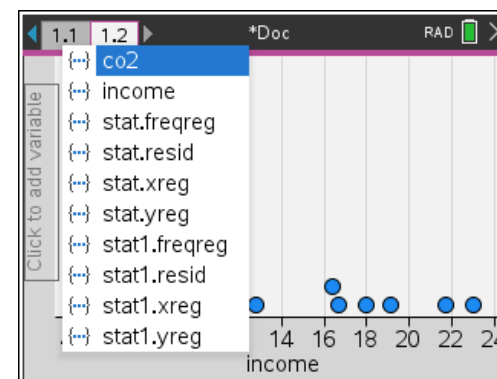
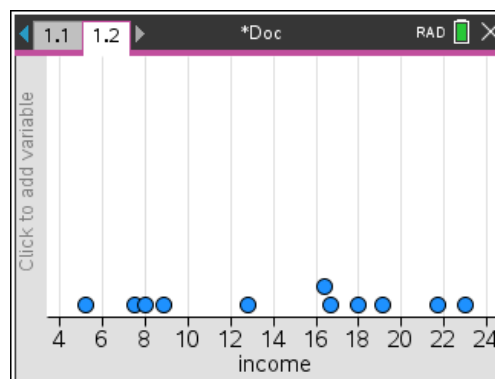
Using the graph drawing capabilities

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



Examples have been extracted, with permission, from the Cambridge General Mathematics Units 1 and 2 Textbook

www.maffsguru.com

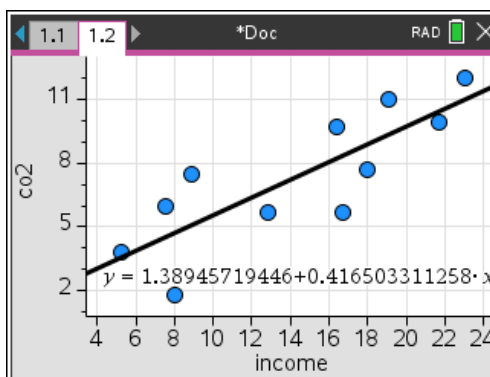
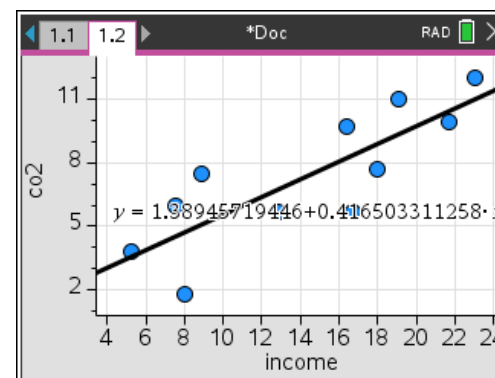
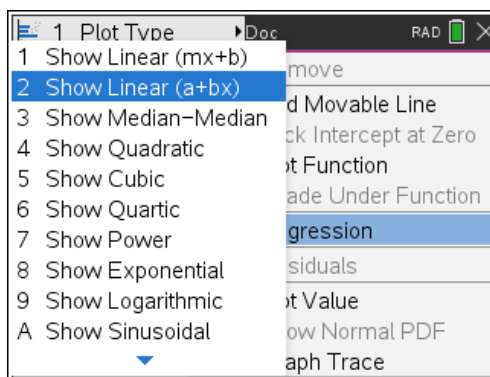
Using the graph drawing capabilities

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 CO₂

Examples have been extracted, with permission, from the Cambridge General Mathematics Units 1 and 2 Textbook

www.maffsguru.com

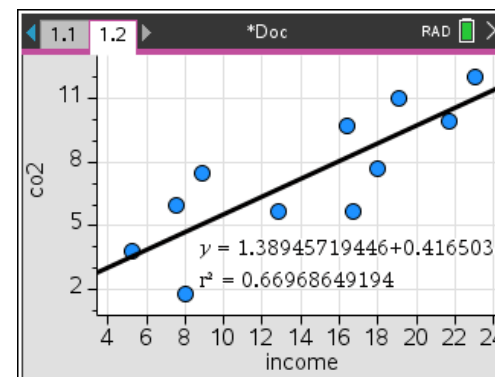
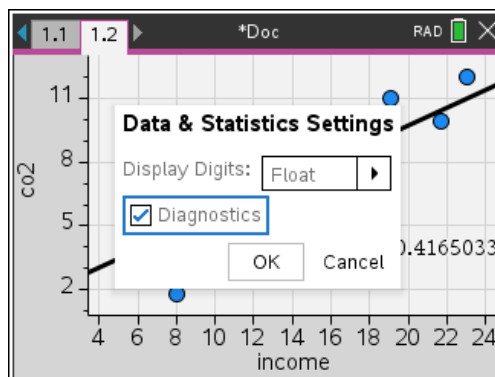
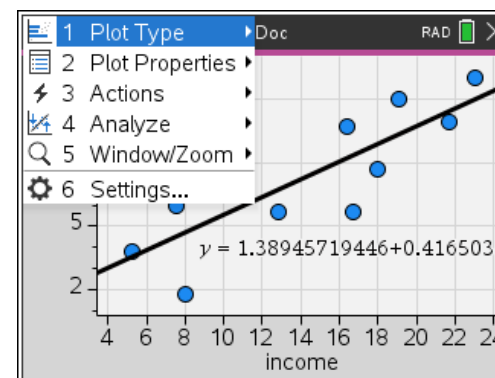
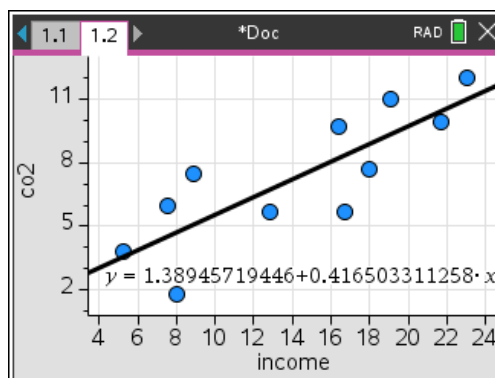
Finding the 'r' value

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

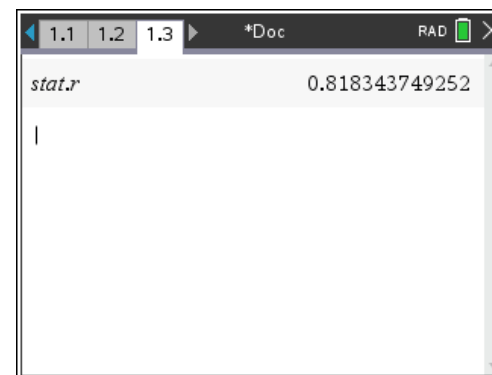
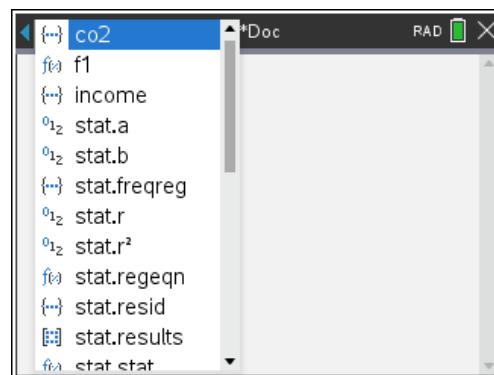
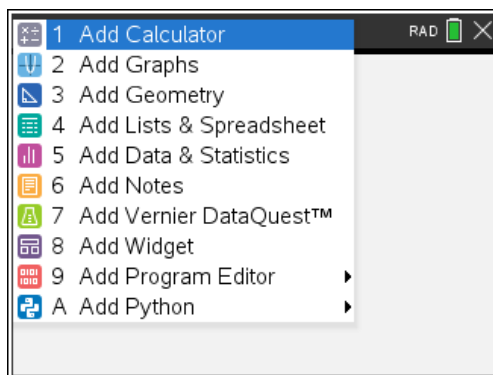


Examples have been extracted, with permission, from the Cambridge General Mathematics Units 1 and 2 Textbook

www.maffsguru.com

Finding the 'r' value

Or ... we can use a CAS function :



Examples have been extracted, with permission, from the Cambridge General Mathematics Units 1 and 2 Textbook

www.maffsguru.com