Using data transformation to linearise a scatter plot

Thursday, 23 January 2020 8:17 pm

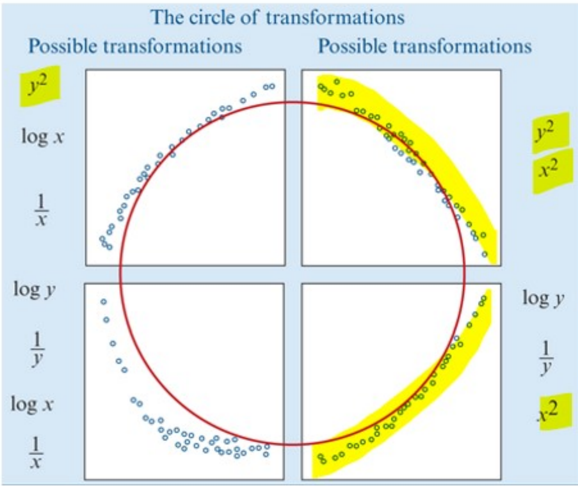
★ By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- How to apply the squared transformation
- Why we need to apply the squared transform

RECAP:

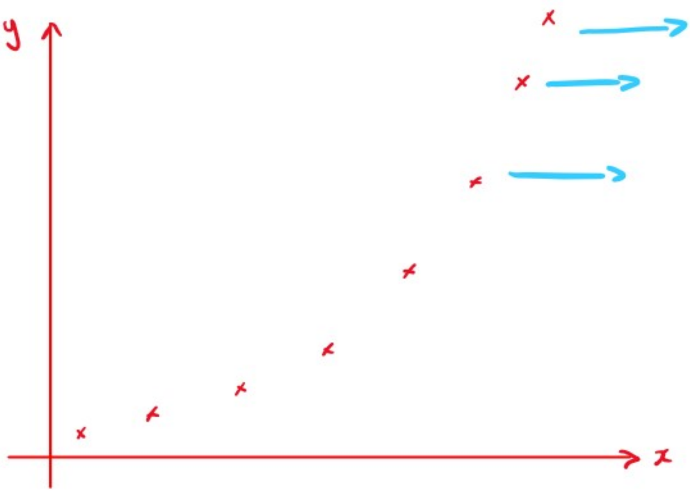
In the last lesson we looked at the fact that not all data is going to be linear.
In fact, lots of it will curve!
When it curves, we need to find ways to make it straight (or straighter) so we can use our linear regression analysis to help us predict results.

The circle of transformations from the Cambridge Further Maths Units 3 and 4 textbook is shown below and we're going to look at the squared transformation in this lesson.



Squaring the x values

When we have data curved as shown below, we are looking for ways in which to be able to stretch the x values to make them straighter.



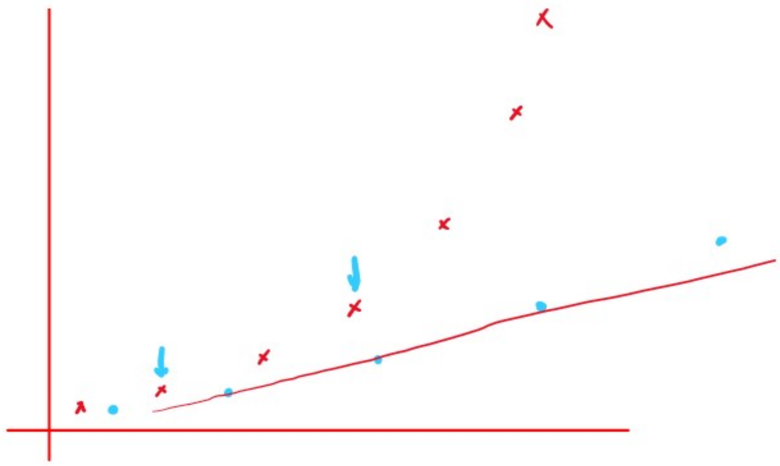
$1^2 = 1$

$2^2 = 4$

$3^2 = 9$

$4^2 = 16$

We can do this by squaring the x-values.
The larger the x-value the more it shifts to the right, hence stretching it.



This is called an x-squared transformation.

In doing this we end up with a graph which is **linear** and hence we can use the equation of a straight line to help us predict values. Although the data is now linear, we need to be careful to go back to something from a previous lesson ...

Label of the x-axis

Label of the y-axis

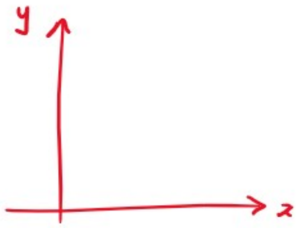
$$y = a + bx$$

y-axis intercept

Gradient of the line

$y = mx + c$

$y = a + bx$



It's **vitally important** that you realise the labels of the x- (and or y-) axes will now change.

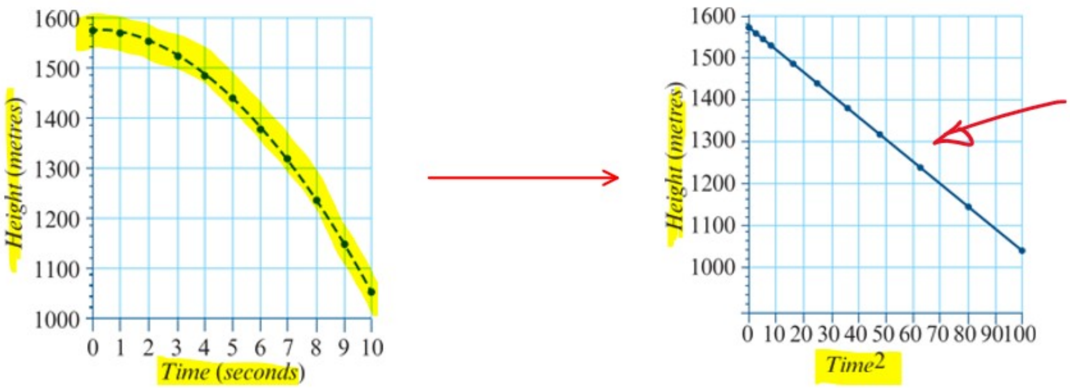
Hence, for the example above, we would have an equation in the form:

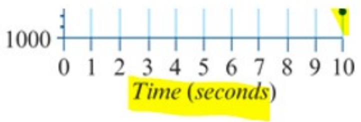
$y = a + b x$

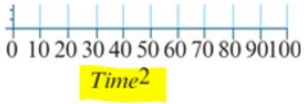
$$y = a + b \times (x^2)$$

The x-axis is now showing all the values of x^2

For a real world example we might have:







$$height = 1560 - 4.90 \times (time)^2$$

y-axis label

Intercept

x-axis label

Gradient:
For every one unit increase in $time^2$ there is a decrease of 4.90 metres in height

An example using the CAS

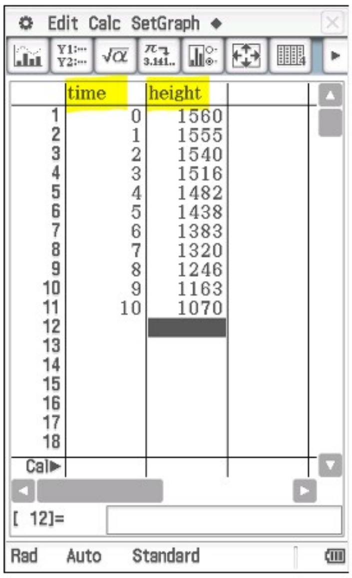
Remember, we can use the CAS in this whole course.
We are used to using the Statistics screen to enter values.
We will enter the raw data items we have and then get the calculator to change the x-values and then plot the appropriate graph.

The following example has been extracted with permission from the Cambridge Further Maths Units 3 and 4 textbook.

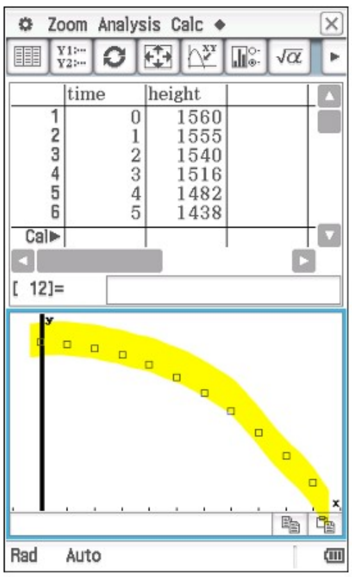
The table shows the height (in m) of a base jumper for the first 10 seconds of her jump.

Time	0	1	2	3	4	5	6	7	8	9	10
Height	1560	1555	1540	1516	1482	1438	1383	1320	1246	1163	1070

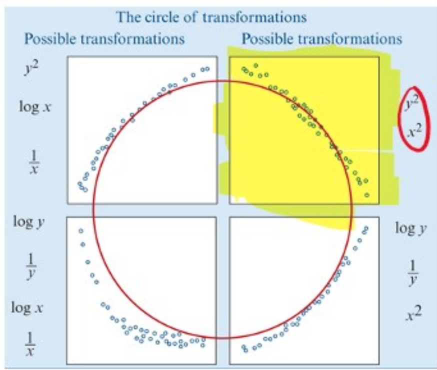
a. Construct a scatterplot displaying height (the RV) against time (the EV).
b. Linearise the scatterplot and fit a least squares line to the transformed data.
c. Use the regression line to predict the height of the base jumper after 3.4 seconds.



Firstly, enter the data as it is shown in the tablet then show the data.

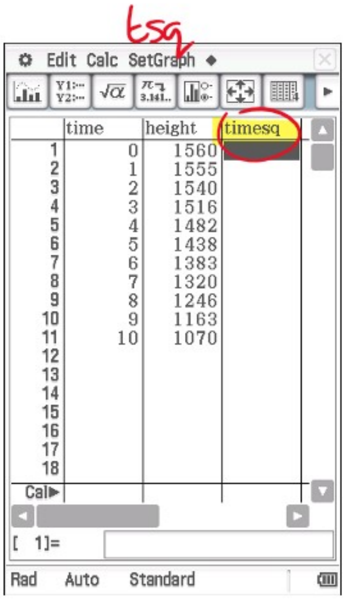


Look at the shape of the data and decide which transform to use using the circle of transformations

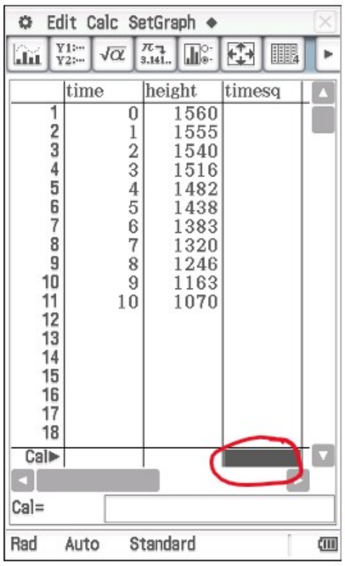


This suggests an x^2 or y^2 transformation.

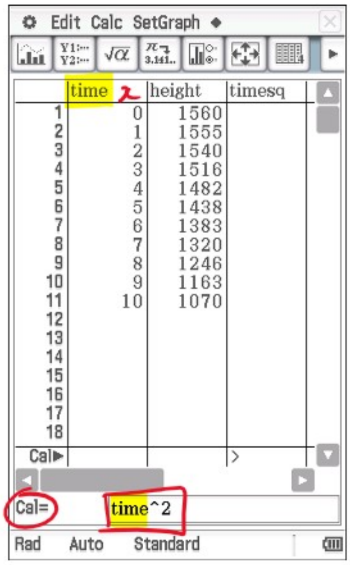
We will use an x^2 transformation



Add a new column called *timesq*.

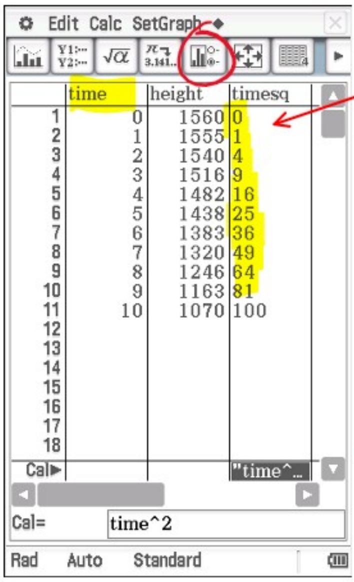


Put cursor in shown cell and type time^2. Your calculator will know what do to.

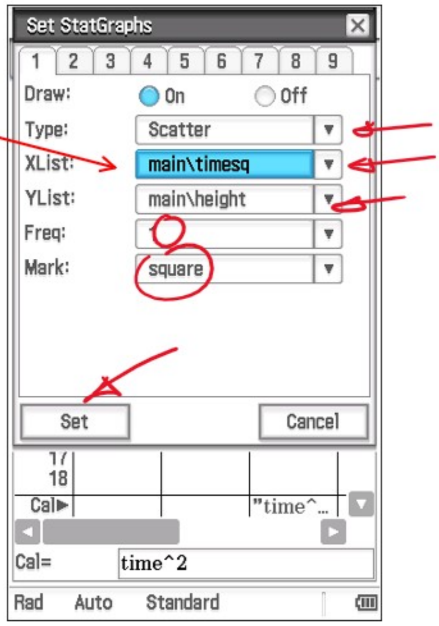


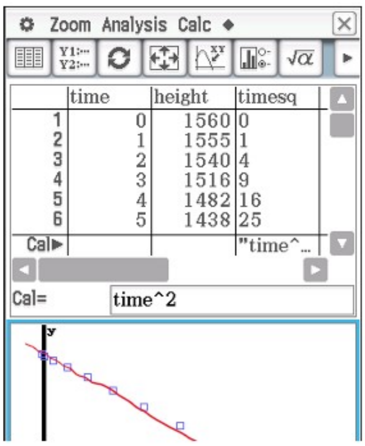
^2

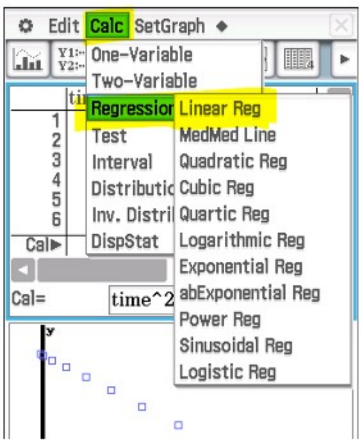
You will notice the formula.
When you hit EXE, you will have new values in this column.

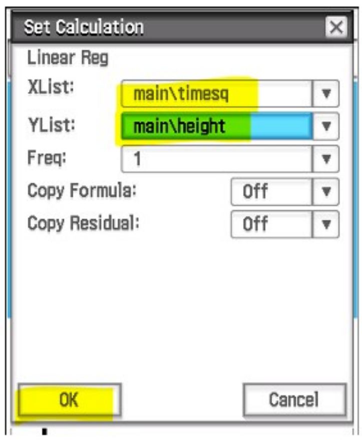


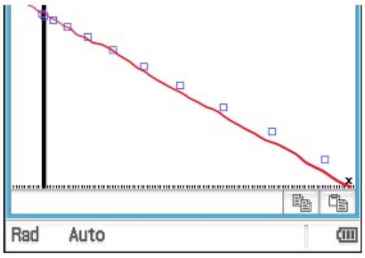
You will now use the new column to draw the scatter plot (and then conduct any regression analysis).

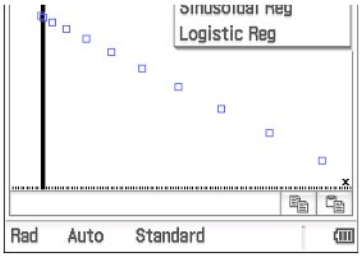


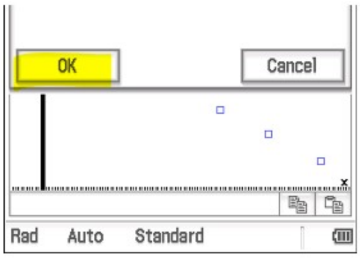






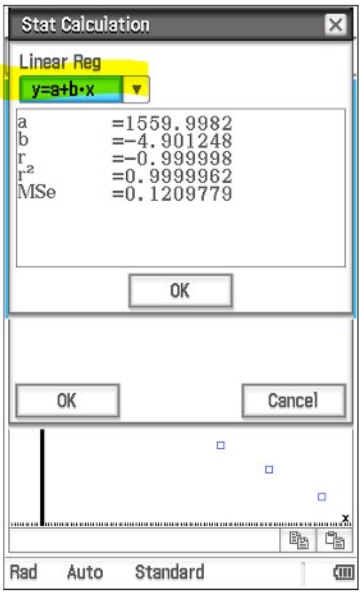






Here is the data which is much straighter than it was. We now extract the data in the same way we did for linear data to enable us to do a regression analysis.

Remember to select *timesq* or you won't get the correct equations.



So, we can now write the linear regression equation as:

$$height = 1560 - 4.90 \times (time)^2$$

$$y = a + bx$$

$$height = 1559.9982 - 4.90(time)^2$$

Making sure you interpolate and extrapolate properly

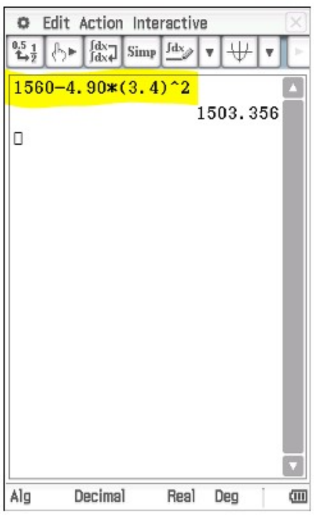
The most common mistake people make now is now using the formula properly.

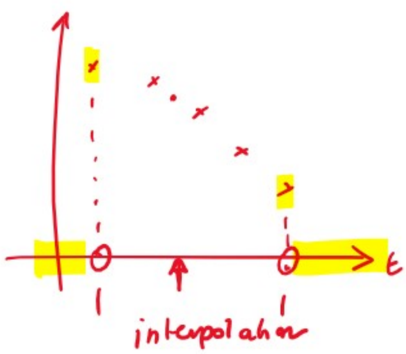
The question asked for the height when the time was 3.4 seconds.

Make sure you substitute the value into your correct equation by **squaring the value of time**.

$$height = 1560 - 4.90 \times (time)^2$$

$$height = 1560 - 4.90 \times (3.4)^2$$





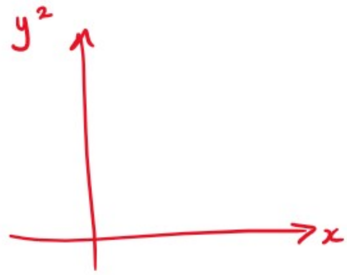
$$height = 1.503\ m$$

Alg Decimal Real Deg

$height = 1,503\ m$

This example showed how to use an x^2 transformation. The same process would be used for a y^2 transformation.

Note: Generally the question will tell you which transform you should use.



$(\text{Height})^2 = 1590 - 4.90 \times \text{time}$

VCAA Exam Question on this concept
2019 Paper 1

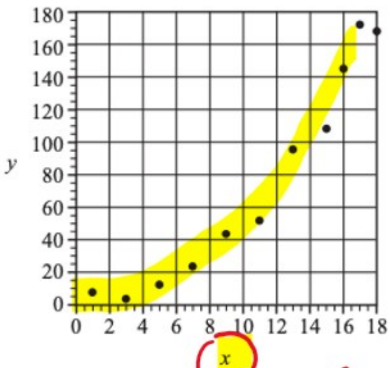
Question 12

The table below shows the values of two variables x and y.

The associated scatterplot is also shown.

The explanatory variable is x.

x	y
1	7.6
3	3.4
5	12.1
7	23.4
9	43.6
11	51.8
13	95.4
15	108
16	145
17	172
18	168



x^2

The scatterplot is non-linear.

A squared transformation applied to the variable x can be used to linearise the scatterplot.

The equation of the least squares line fitted to the linearised data is closest to

A. $y = -1.34 + 0.546x$

B. $y = -1.34 + 0.546x^2$

C. $y = 3.93 - 0.00864x^2$

D. $y = 34.6 - 10.5x$

E. $y = 34.6 - 10.5x^2$