GRADE 12 APPLIED
REGRESSION (DESMOS)

Name:____________
Date: ____________

A graph is very useful way to see what it is that a function does. A mathematical expression is useful too!

It is fairly simple to take a function, evaluate it at some number of data points, plot the data dots, and connect them to make a graph.

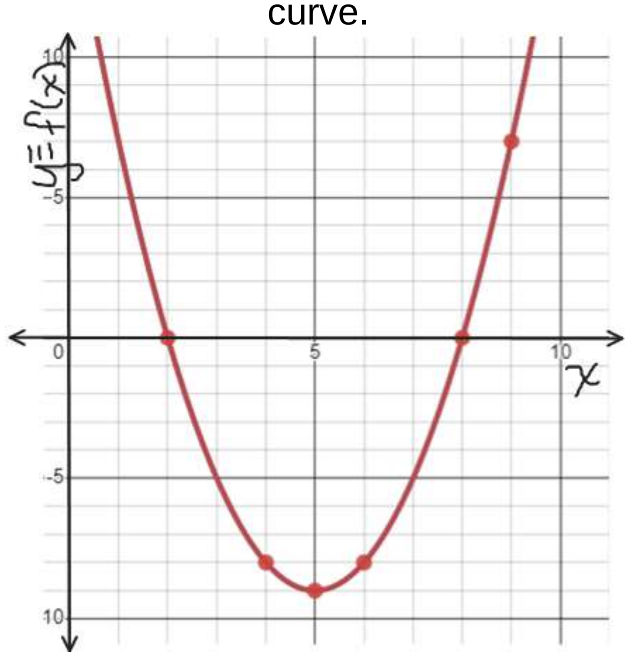
Any decent graphing tool can readily plot data as a 'scatter plot' and draw in the curve that graphs a function. In the old days, arithmetic, pencil , and paper grids were used to do this.

Example: Graph the function. $y = x^2 - 10*x + 16$

Evaluate the function at some values of 'x' to determine the corresponding value of the function so that it can be plotted on a y-axis of a graph.

x	$x^2 - 10*x + 16$	f(x)
2	$(2)^2 - 10(2) + 16$	0
4	$(4)^2 - 10(4) + 16$	-8
5	$(5)^2 - 10(5) + 16$	-9
6	$(6)^2 - 10(6) + 16$	-8
8	$(8)^2 - 10(8) + 16$	0
9	$(9)^2 - 10(9) + 16$	7

The dots are the 'scatter plot' . They are connected with the parabolic curve.



But sometimes all that is known is the data not the function that connects and explains the data. To make it worse, generally any data that is collected has some amount of error in it! Trying to connect the data dots and discover a mathematical function that explains the data is the purpose of REGRESSION.

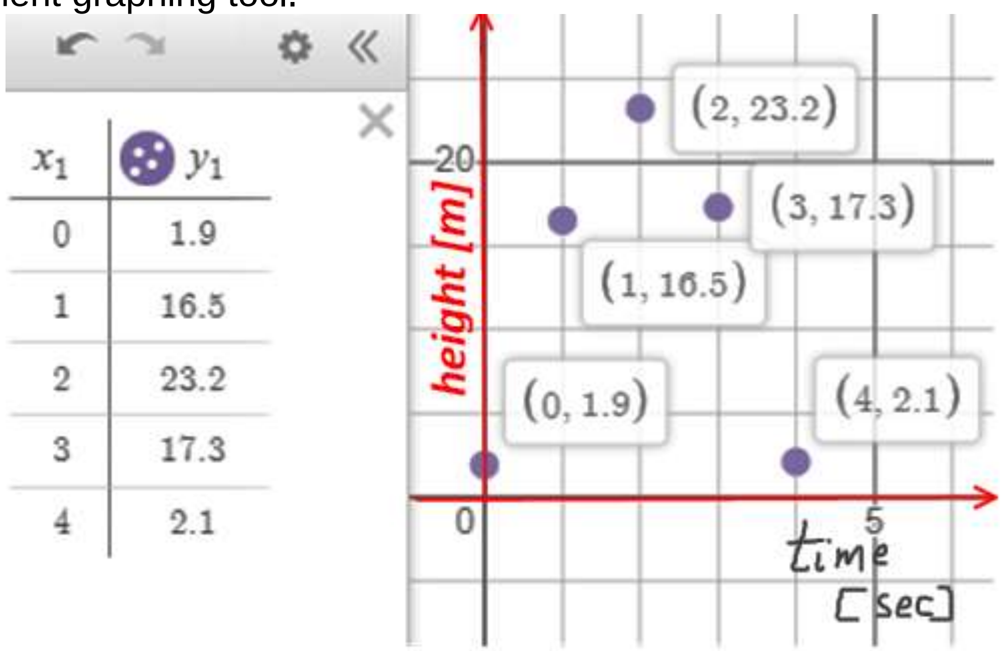
EXAMPLE REGRESSION

A student is making a video of a ball thrown upwards. Using a camera in slow motion the student measured the following data:

- At zero seconds the ball was ~1.9 m above the ground
- At one seconds the ball was ~16.5 m above ground
- At two seconds the ball was ~23.2 m above the ground
- At three seconds the ball was a ~17.3 m above the ground
- At four seconds the ball was ~2.1 m above the ground.

The '~' shows that the measurement is not exact. Anytime a measurement is made there is bound to be, undoubtedly, some amount of statistical measurement error!

The student records the data, plots the dots as a 'scatter plot' using any convenient graphing tool.



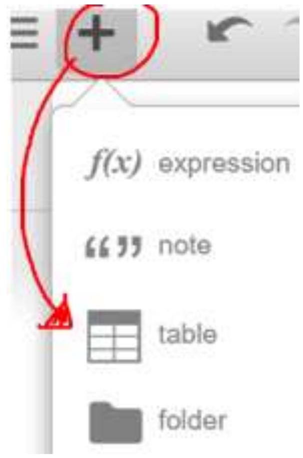
The student wants to find a formula that reasonably explains the data and connects the dots. A formula that expresses the height of the ball **as a function of** the time it is in the air once thrown

Connecting the dots to find a mathematical formula that explains data is called 'REGRESSION'

CALCULATING A REGRESSION (USING DESMOS)

Enter the data into a table, plot the scatter plot dots (as above)

To enter data into a table in DESMOS use the Add Table Feature. →



The student recognizes from the shape of the graph that it is most likely a parabola governed by a quadratic function of some type.

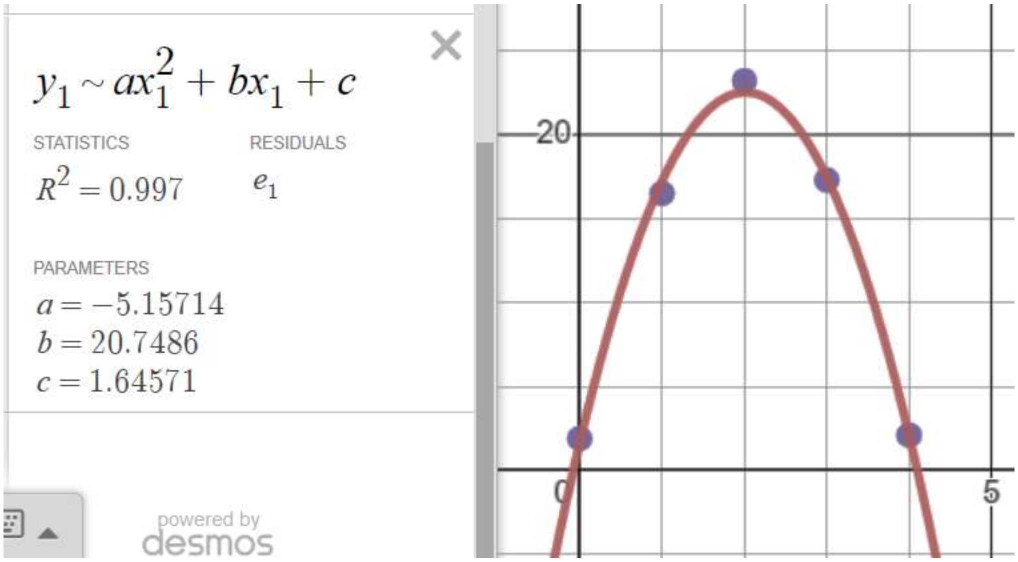
The student adds a general expression for a quadratic function in DESMOS as follows:

$$y_1 \sim ax_1^2 + bx_1 + c;$$

which pretty much means a particular ‘y’ is approximated by some particular x. The ‘~’ is readily available on a laptop keyboard top left, or as a selection in the keyboard selection of the DESMOS App.

The reason we use the ‘~’ , the approximation symbol, is because the formula will not be exact since the data is not exact, but it will be a statistically *good fit*.

The result from DESMOS is the following:



Notice the data points on the scatter plot are not *perfectly* connected! That is because the data was likely not perfect, seldom is any data measurement perfect anyway. But the parabolic curve does a pretty good job of fitting as best as possible to the dots in the scatter plot.

The gobbly gook expression at the left shows that the formula that best fits a quadratic function to the data is given by:

$$\boldsymbol{y = -5.15714x^2 + 20.7486x + 1.64571}$$

So, this formula can be reliably used to predict the height of the ball if you do the experiment again!

When presenting your solution to someone you would likely round off some of the coefficients in the formula to:

$$\boldsymbol{y = -5.16x^2 + 20.75x + 1.65,}$$

But, of course, by rounding you are re-introducing some error.

How reliable the formula is given by the R^2 value, the concept of a 'correlation coefficient': how well the formula correlates and predicts the data. A value for R^2 of 99.7% *suggests* that the calculated formula is *likely* very reliable. A complete discussion however of the R^2 factor is well beyond the scope of High School math

YOU TRY

YOU try all the above, ensure you get the same result!

EXAMPLE – FIT AN EXPONENTIAL FUNCTION

Perhaps you know that bacteria are a life form just like you and I. They reproduce. You are a scientist counting bacteria in a dish in an experiment. You want to derive a formula which predicts how the number of bacteria increase as a *function of time*.

Here is table of your data, number of bacteria vs time in hours.

Time [hours]	0	1	2	3	4
Nbr of bacteria	16	34	60	133	250

Steps:

- **Add the table** of measured data into DESMOS. The independent variable, time in this case, as the x_1's, and the dependent variable, number of bacteria that you are counting, in the 'y_1'. Generally 4 or 5 points is sufficient; well, at least two for a line, at least three for a quadratic, etc), but the more data entered the better the result.

- Note the scatter plot that DESMOS automatically makes.

- Determine from the shape of the scatter plot which function might best fit the data that is entered. A line, a parabola with one hump, a cubic with two humps, an exponential, etc. This particular data appears to be exponential, just like compound interest! So $y_1 \sim a * b^{x_1}$ is the general form of an exponential function

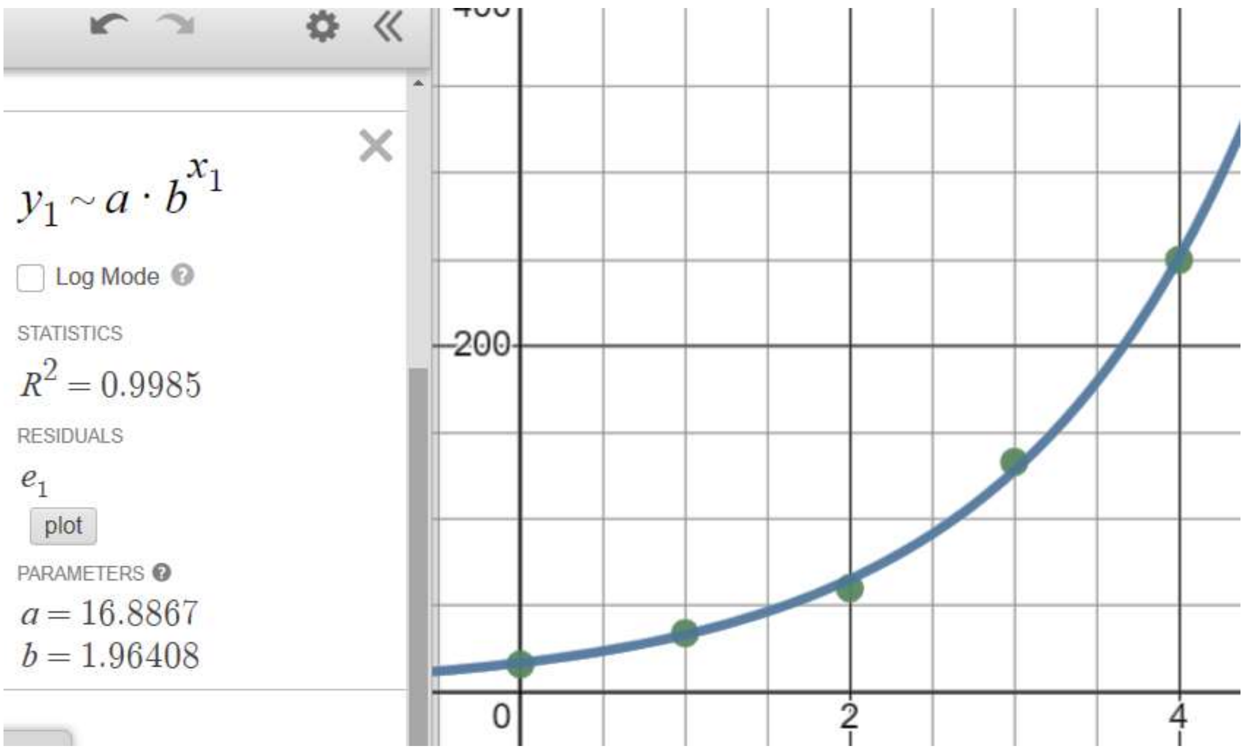
- Perform the regression using the ' **y~** ' and the general form of the formula you are trying to fit. example: $a*x_1^2 + b*x_1 + c$ for a general quadratic.

- Note the calculated coefficients in front of the different x terms and any constant at the end.

- Note the correlation coefficient R^2 to see how notionally reliable the calculation is

- Record the function that you have calculated.

Here is what you should have got:



So the expression you should have got is that the number of bacteria, y, grow as a function of time, x, as follows:

$$y = 16.8867 * 1.96408^x$$

It seems to have a strong correlation coefficient, 99.85%, so it is likely accurate (at least for the domain we studied so far , the first 4 hours)

CONCLUSION

Being able to derive a reliable mathematical expression to explain different data is a very important way to model the real observed world.

Regressions are rather important