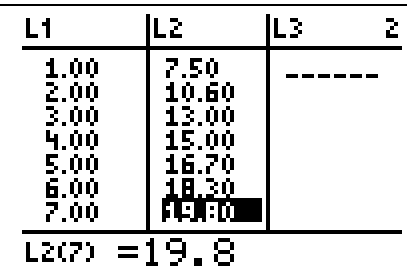
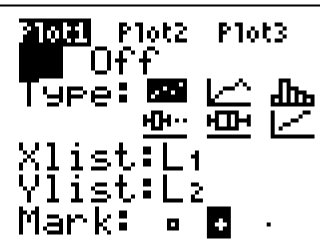
Regression-two or more variables related.

So what is this “regression”? Does it mean that you’ll forget most of what you’ve learned in previous sessions by the time you finish this one? No, it’s functionally relating a dependent variable to one or more independent variables, typically with “best fit” algorithms. The TI83 has several forms for one independent variable. You may explore multiple regression in Minitab. In multiple regression the dependent variable is driven by more than one independent variable. Here we will take and dodge the in and outs of regression in one independent variable.

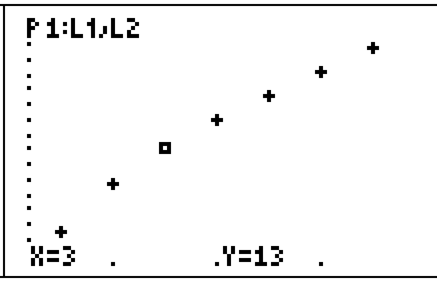
Sometimes we really don’t need regression because the functional relationship is already known. It is edifying to do regression anyway in such cases to see how it could be applied when the relationship is not known or not determinable by other means. Suppose you have a spring tube that launches steel balls horizontally off a wall. You have a wooden floor, which yields a small pockmark when struck with this steel ball. You hold the tube at distances 1, 2, 3, 4, 5, 6, and 7 feet from the floor and fire the steel ball. The corresponding pock marks on the gym floor measure 7.5, 10.6, 13.0, 15.0, 16.7, 18.3, and 19.8 feet from the wall. What is the velocity (speed) of the steel ball and what is the distance (horizontal) as a function of height? If you obtain a ladder and launch from 12 feet up the wall, how far across the floor will the ball strike assuming the coach hasn’t caught you yet and had you expelled and escorted off campus? Of course, you could easily answer all these questions from your knowledge of simple physics, but

First, make a **scattergram** of the data.

STAT-1. Place the distances in L2 and the heights in L1.	

2nd-STATPLOT-1. Select *On* and the first Type. Make *Xlist* L1 and *Ylist* L2. Select a mark.	

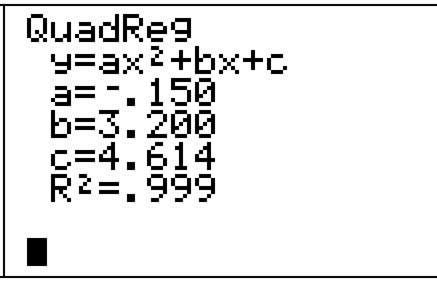
ZOOM-9-**TRACE**.



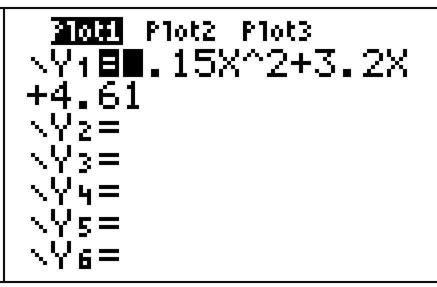
The data displays subtly concave down but could be mistaken for linear. You assume a curve and someone says "Do a quadratic regression- that's non-linear". So you do a quadratic regression with the height up the wall as the independent or driving variable and the distance along the floor as the dependent or driven variable:

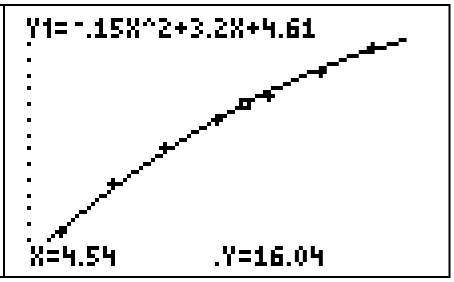
STAT-*CALC*-*5*-(key in "(**2nd-L1**,**2nd-L2, VARS**-*Y-Vars*-*1*)" Actually, you could do without the second parenthesis.

ENTER. If you did not get the R^2, do **2nd-CATALOG-D-**and select *DiagnosticOn* and **ENTER** and **ENTER** again to receive "Done" on the home screen. Then redo the regression. An R^2 close to one means a very good fit.



The regression function is saved in Y1. Check with **Y=.** Note that *Y1* is automatically enabled and that *Plot1* is still enabled. We want both enabled. If either is not, cursor to it and **ENTER**.



GRAPH-TRACE or **TRACE.** Window settings of the previous **ZOOM-***9* were saved and both the plot and the regression function are graphed, showing the regression to be a good fit for the data. You can cursor up and down to switch between the scattergram and function for trace.	

So now we have this regression function in Y1. What good is it? Well, for one thing you can extrapolate or predict values outside the data range. For example, for a height of 12 feet up the wall you can evaluate the function at X=12:

VARS-*YVARS*-1-**ENTER** - finish as shown-**ENTER**	$Y_1(12)$ 21.410

Can we depend on this prediction? Since we know all about the physics behind this kind of action, let's figure it out.

As the steel ball falls from rest, the distance D in feet is $16.1*t^2$, t in seconds. At the same time it is falling, it is progressing horizontally at rate V feet per second so that the horizontal position is V*t. Solving both equations for t and setting these solutions equal to each other yields the relationship:

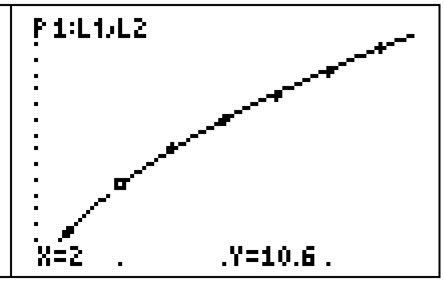
$$D = 0.249*V*\sqrt{H}$$

V must be measured and actually we've already done it! Take a data point that falls close to the function graph as all do in this case, say (4, 15.0). Plug in the 4 for H and the 15.0 for D to get the final relationship and even the value of V.

V=30.12 feet per second and $D = 7.5\sqrt{H}$

So what do we predict for a height of 12 with this function of H=12? 26 feet, not too close at all to the regression prediction.

Now do a another regression on L1 and L2. Repeat the above exercise except under **STAT-***CALC* select *A*, *PwrReg*, the power regression instead of *5*, *QuadReg* like you did before.

PwrReg (L1,L2,Y1)	PwrReg y=a*x^b a=7.50608 b=.49814 r²=.99998 r=.99999	

The fit for this function looks good, too. How is it at extrapolation outside the data range? Y1(12) for this function gives 25.9, a lot closer to the actual 26 than the quadratic yield. So which form of regression do you trust to be the most likely correct one here? Of course you noticed that the power regression has the same form as the physics derived relationship.

A man goes on a hike each week. Like a bouncing ball, he retains only a certain percentage of his energy from week to week (he's getting old and lazy, perhaps) so that his distance traveled each week is a set percent (less than 100) of the previous week's distance. Place 1, 2, 3, 4, 5, 6, 7 in L1 to be the hike number and 6.0, 4.8, 3.8, 3.1, 2.5, 2.0, 1.6 in L2 for the corresponding distances in miles. We want a function that yields list 2, the dependent variable , from list 1, the independent variable. Do an ExpReg:

STAT-*CALC*-0 and finish with "L1,L2,Y1"- **ENTER** as before;	ExpReg y=a*b^x a=7.43261 b=.80324 r²=.99981 r=-.99990

So you could scattergram this and graph it with the regression in Y1 as before. What distance would this function predict for the 12th hike? Do you remember how to retrieve Y1 from **VARS**? For a hike before the first one recorded? 0.54 miles. 7.43 miles. Hint: The last answer is Y1(0). Try Y1(-1). What does this answer represent?