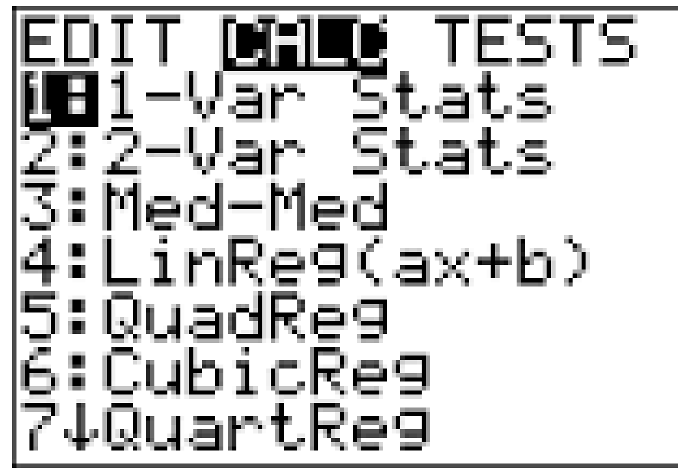
Regression for non-linear data
don’t let students go model shopping

NCTM National Conference
Boston MA
April 2015

Julie Graves
graves@ncssm.edu
The North Carolina School of Science and Mathematics

Model shopping….

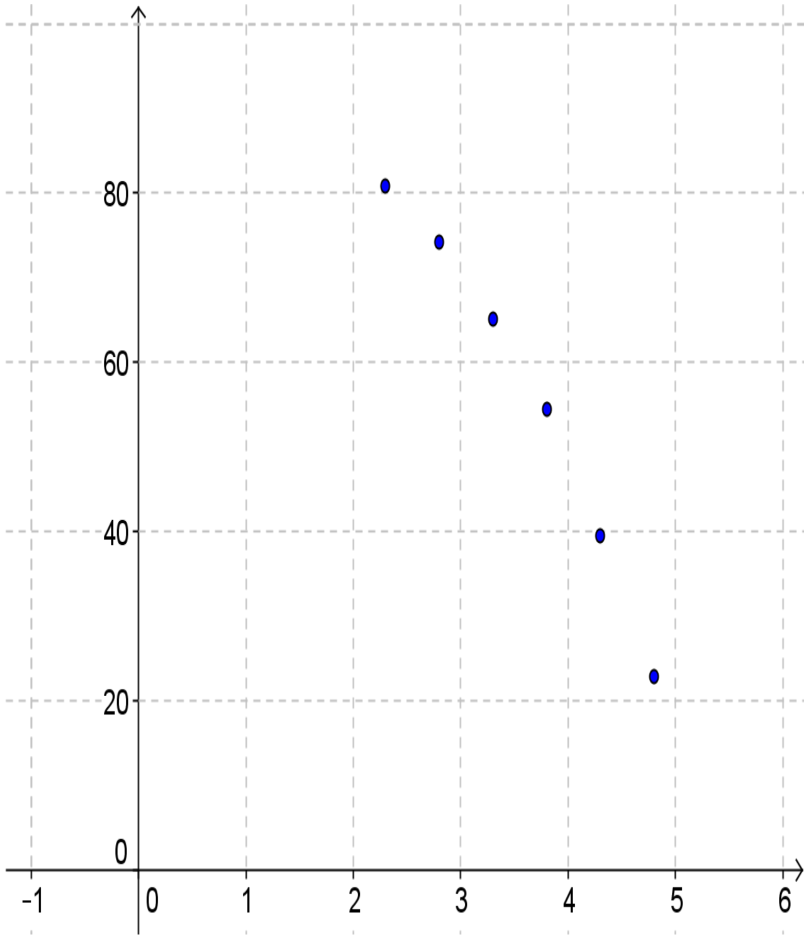
- …refers to the practice of "trying on" all of the regression choices available on the calculator, and then choosing the model that "looks the best".
- This is not a reliable way to choose a model and does not develop understanding about functions or transformations.



Ball Thrown from a Roof Top

We will find a model for data gathered by throwing a ball from the roof top of a building and recording its height above ground and the time that has elapsed since it was thrown. Time is measured in seconds and height is measured in feet.

Ball from a Roof Top: Height versus time



(2.3, 80.8)
(2.8, 74.2)
(3.3, 65.1)
(3.8, 53.5)
(4.3, 39.5)
(4.8, 22.9)

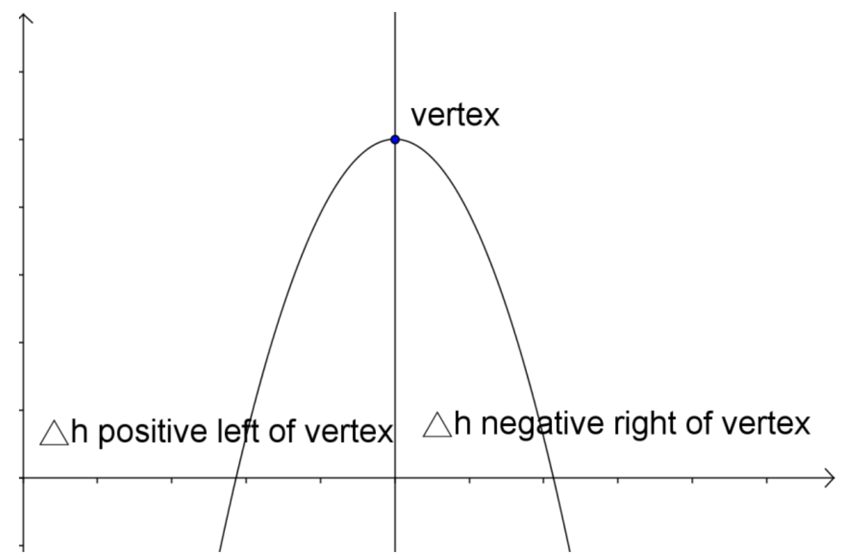
Where is the vertex???

The curvature of the scatterplot suggests a downward opening parabola. The scatterplot does not show the location of the vertex, but we can study the data to find the *t*- coordinate of the vertex.

The data will show where the vertex is located.

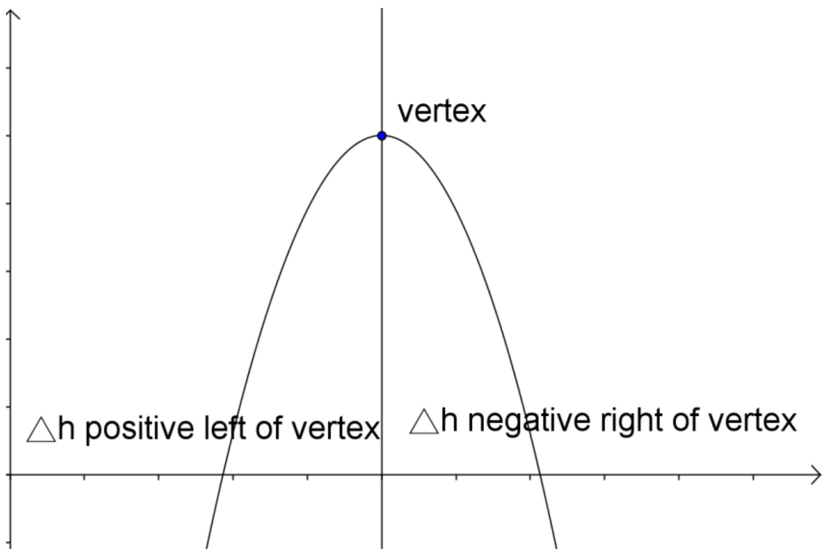
Note that for any downward opening parabola, the h-values are increasing to the left of the vertex and are decreasing to the right of the vertex.

Thus, Δh values change from positive to negative at the vertex.



We can find the t-coordinate of the vertex by finding the t-value at which Δh would change sign.

Ordered pairs of the form $(t, \Delta h)$ will allow us to find this t-value.



Using the TI-84 list operation Δ `list`, we can efficiently find values of Δh.

Since all of our data points are to the right of the vertex, all of the values of Δh that we have found are negative.

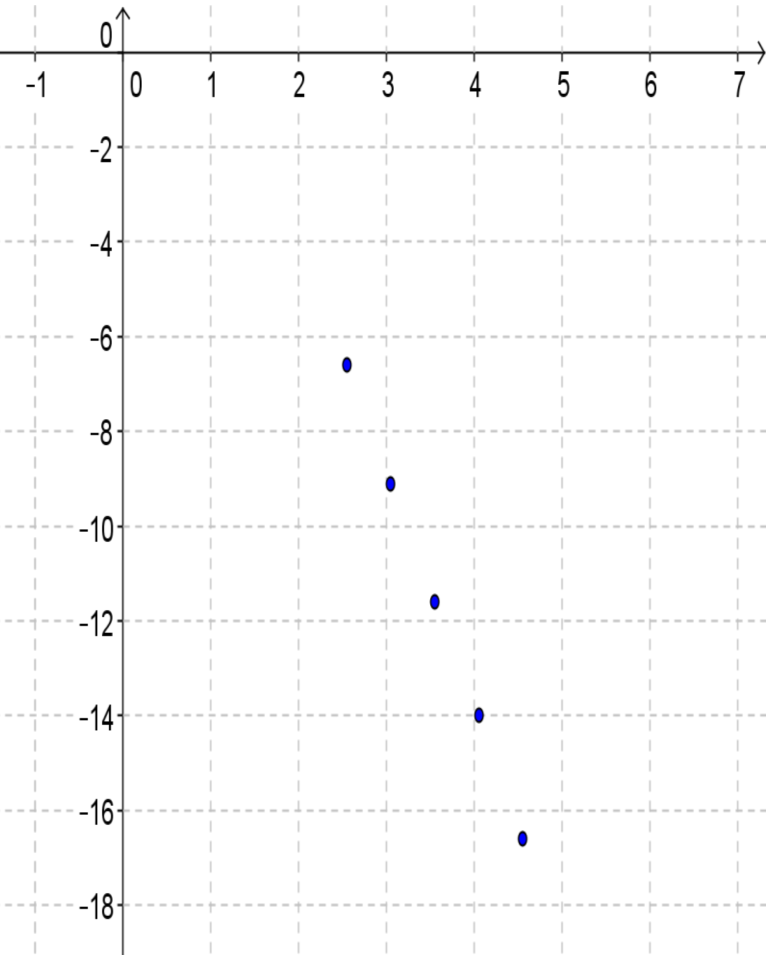
L1	L2	L3	L4
2.3	8Ø.8		-6.6
2.8	74.2		-9.1
3.3	65.1		-11.6
3.8	53.5		-14
4.3	39.5		-16.6
4.8	22.9		

L1	L2	L3	L4
2.3	8Ø.8	2.55	-6.6
2.8	74.2	3.Ø5	-9.1
3.3	65.1	3.55	-11.6
3.8	53.5	4.Ø5	-14
4.3	39.5	4.55	-16.6
4.8	22.9		

Note that we have 6 values for t and 5 values for Δh.

To remedy this mismatch, we can find the average of each pair of consecutive t-values.

Thus, the ordered pairs (2.3, 80.8) and (2.8, 74.2) yield the pair (2.55, -6.6).



The scatterplot of ordered pairs (t, Δh) shows a linear relationship.

The equation of the least squares line is

$$\Delta h = -4.98t + 6.099$$

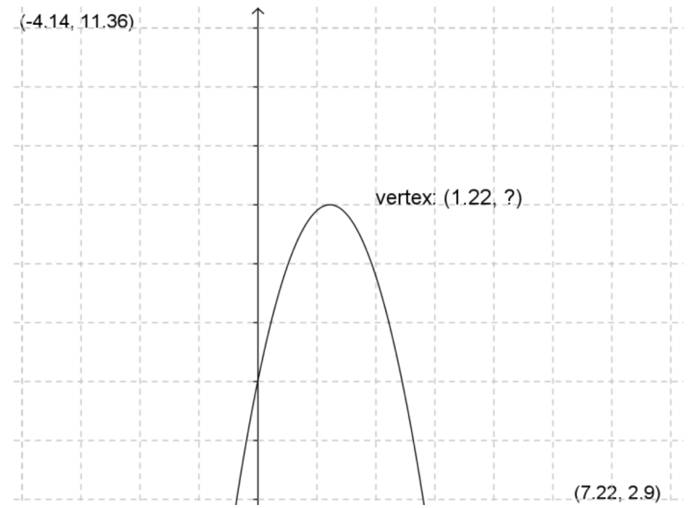
A residual plot would confirm that this linear equation is an appropriate model for these ordered pairs and does indeed successfully represent these data.

The equation $\Delta h = -4.98t + 6.099$

means that Δh would be zero when

t has the value $\frac{6.099}{4.98} \approx 1.22$

The data has shown us that the t-coordinate of the vertex of a quadratic model is about 1.22. This implies that the quadratic model has an equation of the form



$$h = a(t - 1.22)^2 + k$$

Linear Regression to find Quadratic Model

The equation $h = a(t - 1.22)^2 + k$
models a quadratic relationship between t and h.
But it also implies that there is a **linear** relationship between $(t - 1.22)^2$ and h.

We can find appropriate values for each of the constants a and k using linear regression on ordered pairs $\left((t - 1.22)^2, h\right)$

Linear Regression to find a Quadratic Model

L1	L2	L3	L4	L5
2.3	80.8	2.55	-6.6	1.1664
2.8	74.2	3.Ø5	-9.1	2.4964
3.3	65.1	3.55	-11.6	4.3264
3.8	53.5	4.Ø5	-14	6.6564
4.3	39.5	4.55	-16.6	9.4864
4.8	22.9			12.8164

The entries in L5 are

the squares of L1 values minus 1.22.

$$L5 = (L1 - 1.22)^2$$

Quadratic Model

Linear regression on ordered pairs $\left((t-1.22)^2, h\right)$ yields a slope of -4.97 and an intercept of 86.60.
So $\mathrm{a} = -4.97$ and $\mathrm{k} = 86.60$

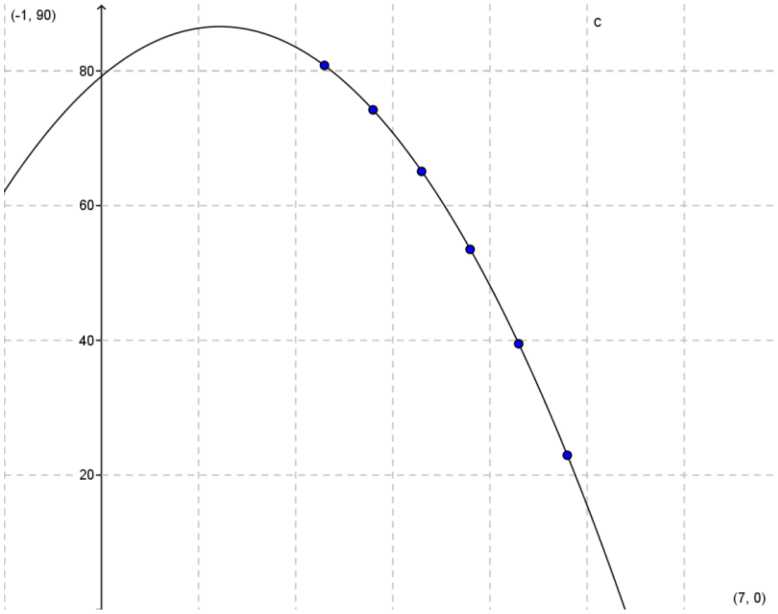
These values for a and k allow us to complete our quadratic model:

$$h = -4.97(t - 1.22)^2 + 86.60$$

Our Quadratic Model Looks Great with the Data

A residual plot would confirm that the data are well represented by

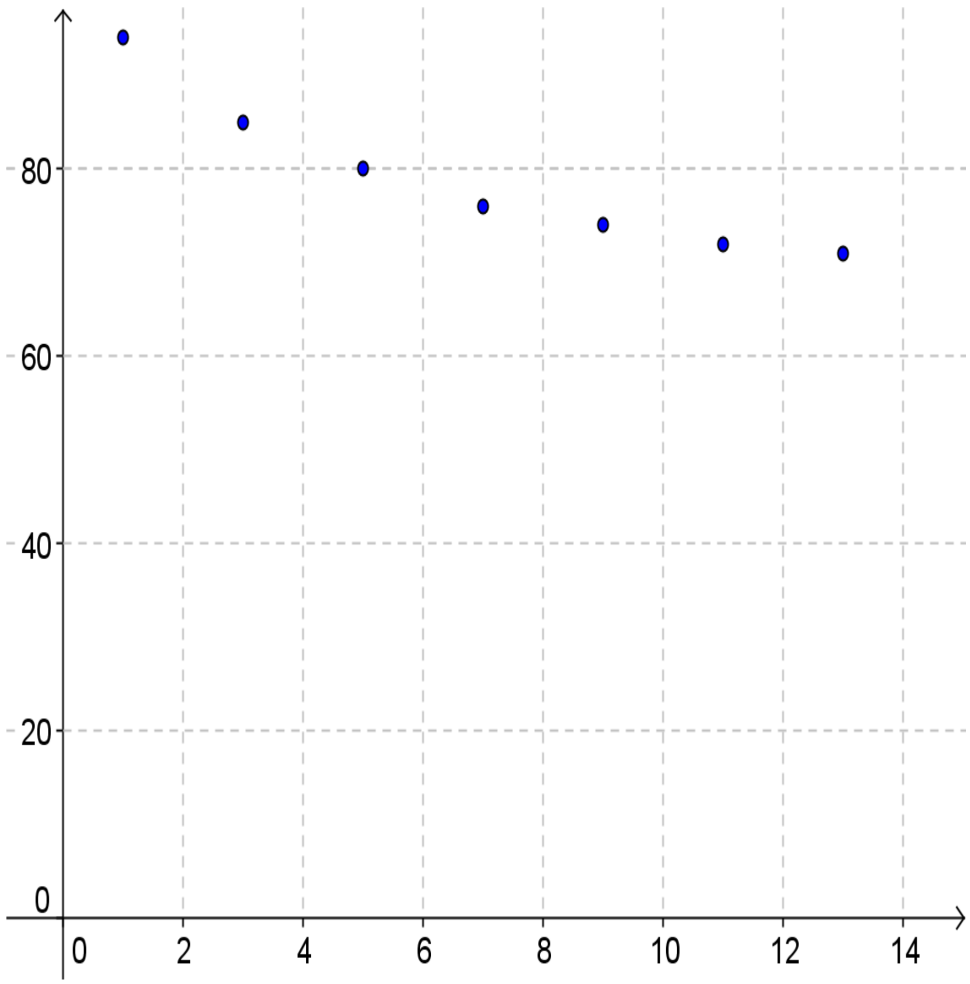
$$h = -4.97(t - 1.22)^2 + 86.60$$



Cooling Coffee

We will find a model for data gathered by measuring the temperature of a cup of hot coffee that cools down until it reaches room temperature. Time is measured in minutes and temperature is in degrees.

Cooling Coffee: Temperature versus Time



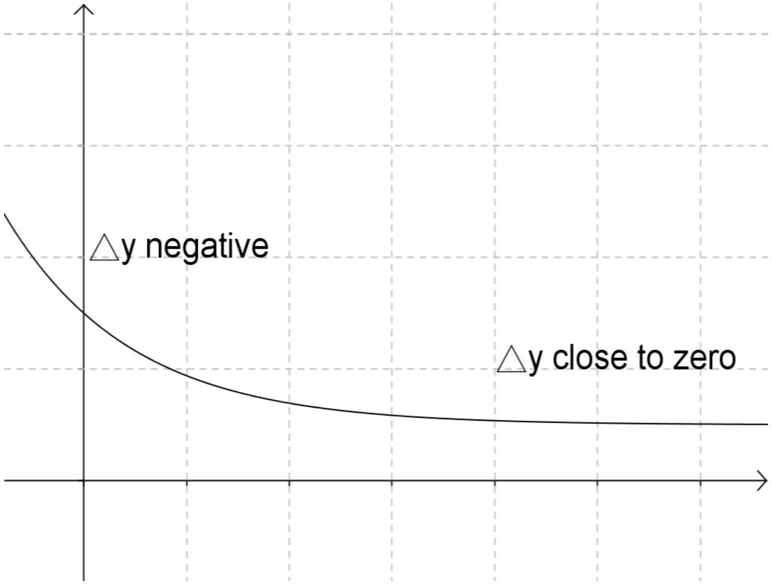
(1,94)
(3,85)
(5,80)
(7,76)
(9,74)
(11,72)
(13,71)

The data want an exponential model

The scatterplot and theory both suggest that this data requires an exponential model. Such a model will have a horizontal asymptote. It is pretty certain that for this data set the horizontal asymptote is NOT at $y=0$. Yet the only exponential models that the TI84 is programmed to find are those whose graphs approach the horizontal axis asymptotically.

Where is the horizontal asymptote?

- For a decreasing exponential model, the y-values decrease and so Δy is negative.
- The y coordinate of the horizontal asymptote represents the y value at which y values stop changing.
- We will use the data to predict the y value at which Δy would be zero.



We can again use the TI-84 list operation Δ `list` to find values of Δy.

We are interested in finding the y-value at which Δy would be zero.

L1	L2	L3	L4	
1	94		-9	
3	85		-5	
5	80		-4	
7	76		-2	
9	74		-2	
11	72		-1	
13	71			

The table shows 7 values for y in L2 and 6 values for Δy in L4. Since there is a dimension mis-match, we can pair each value of Δy with the average of the corresponding consecutive y values. These are shown in L3.

L1	L2	L3	L4
1	94	89.5	-9
3	85	82.5	-5
5	80	78	-4
7	76	75	-2
9	74	73	-2
11	72	71.5	-1
13	71		

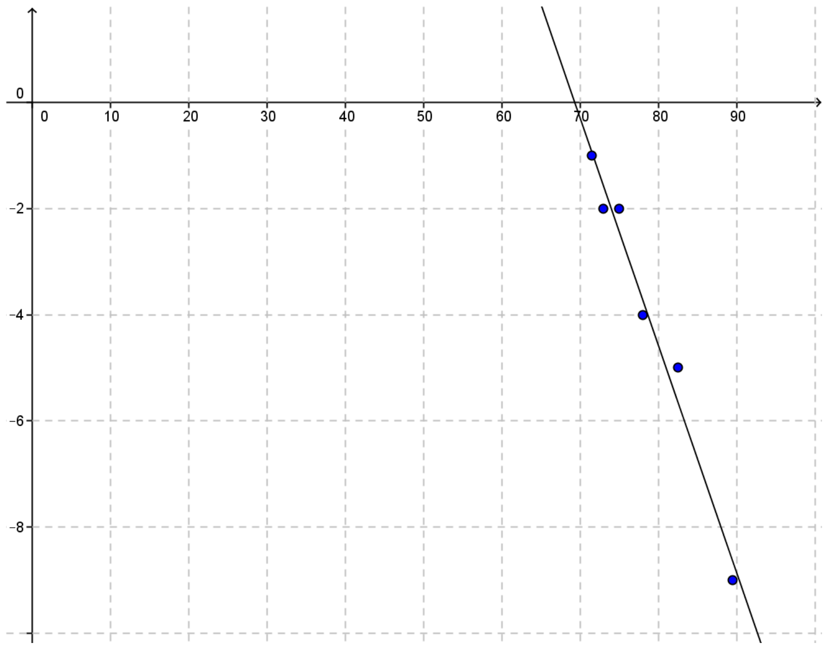
The scatterplot of ordered pairs

(y, Δy) shows a linear relationship. The equation of the least squares line is

$$\Delta y = -0.428y + 29.660$$

This means that Δy would be zero when y has the value

$$\frac{29.660}{0.428} \approx 69.3$$

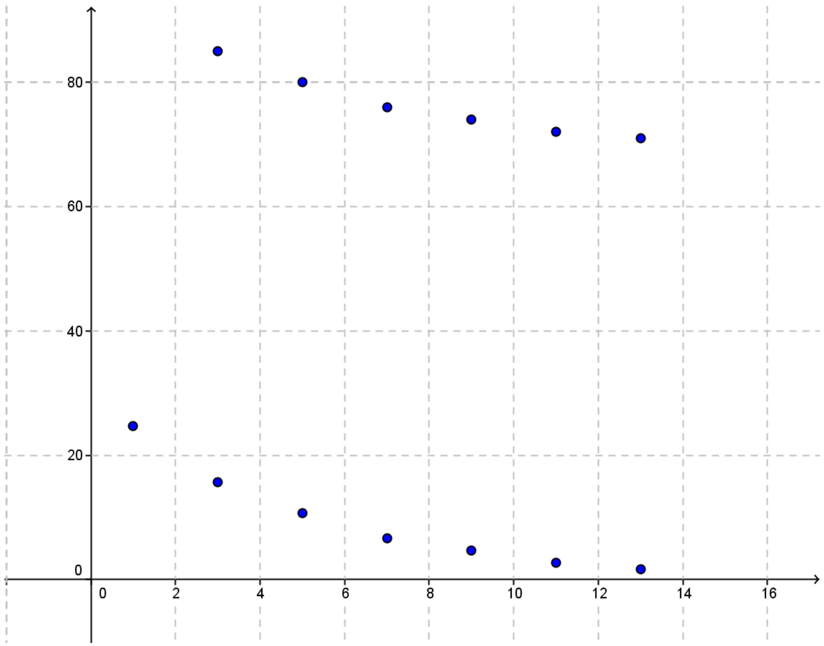


A shifted exponential model

- The data have told us that the exponential model has a horizontal asymptote at $y = 69.3$.
- We are looking for an exponential equation of the form $y = a * b^x + 69.3$

- The TI84 regression options will let us find exponential models of the form $y = a * b^x$. Equations of this form have a horizontal asymptote at $y = 0$.

If we subtract 69.3 from the y-coordinate of each of our original data pairs, the resulting points will be represented by a model that approaches $y = 0$ as an asymptote.



The values in L3 are the values in L2 minus 69.3.

When the TI84 does exponential regression on L1 and L3 , the result is

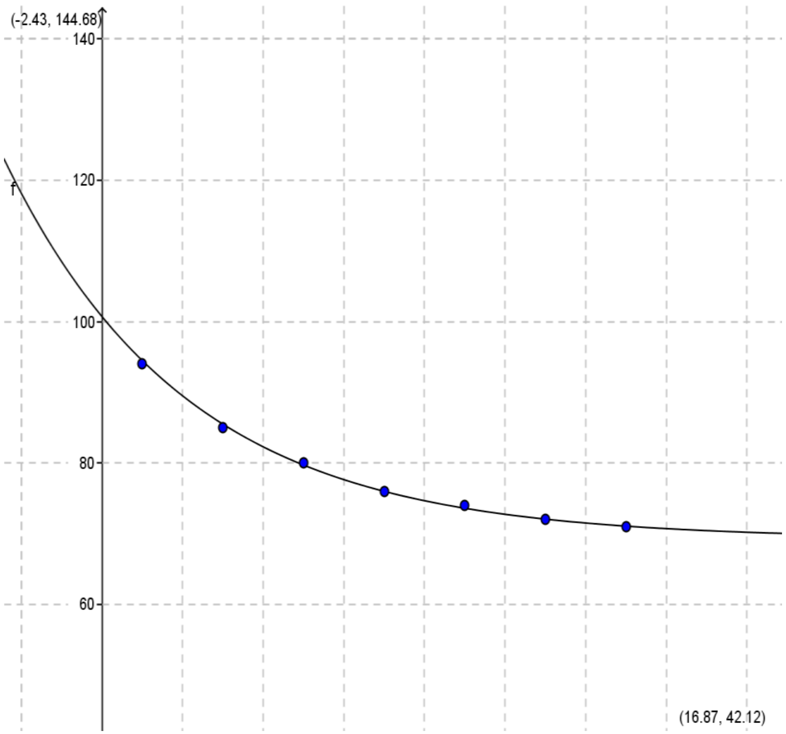
$y = 31.41(.802)^x$.

Since the values in L3 are actually values of y-69.3, we have

$y - 69.3 = 31.4(.802)^x$ or

$y = 31.4(.802)^x + 69.3$

L1	L2	L3
1	94	24.7
3	85	15.7
5	80	1Ø.7
7	76	6.7
9	74	4.7
11	72	2.7
13	71	1.7



The graph show the model

$$y = 31.4(.802)^{x}+69.3$$

superimposed on the scatterplot of the data.

The model implies that the coffee is loosing about 20% of its temperature every minute and that it is cooling in a room whose temperature is about 69 degrees.

Thank you!

Please send me email if you have questions.

graves@ncssm.edu

NCSSM

Job Opportunities at NCSSM

http://www.oshr.nc.gov/jobs/.

Search Job Positions then

Search Departments - NCSSM

Conferences

1. Bridges Conference: Mathematics, Music, Art, Architecture Culture
   - Baltimore, MD July 29 - Aug 1, 2015

2. Anya Greer Math, Science and Technology Conference
   - Phillips Exeter Academy, June 21 - 26, 2015

3. MAA/PREP Teaching Mathematical Modeling as
   - Creating Mathematical Discovery
   - Lincoln, NB July 20 - 25

NCSSM Teaching Contemporary Mathematics Conference

- The NC School of Science and Mathematics
- Durham, NC, January 29 – 30, 2016
- http://www.ncssm.edu/courses/math/tcm/TCM2015/

NCSSM

NCSSM

MATHEMATICAL MODEL BUILDING

Mathematics is studied because it is a rich and interesting discipline, because it provides a set of ideas and tools that are effective in solving problems which arise in other fields, and because it provides concepts useful in theoretical studies in other fields.

From Indian University *http://www.indiana.edu/~hmathmod/modelmodel.html*