

Three Ways to Re-express Quadratic Data (corrected)

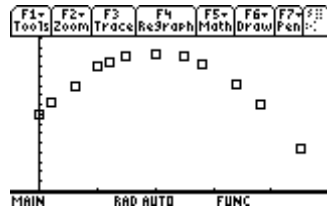
Peggy Craft, NCSSM

TCM Conference February 2004

Here's hypothetical data for a softball thrown from the roof of a building, which we will denote by (x, y) :

time (sec)	0	0.2	0.6	1.0	1.2	1.5	2.0	2.5	2.8	3.4	3.8	4.5
ht (ft)	80	92	110	130	134	142	144	140	132	112	90	44

and the scatterplot¹ of (x, y) :



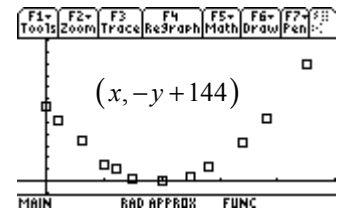
It appears that the vertex is at $(2, 144)$.

Reexpression technique #1: "undoing the y's"

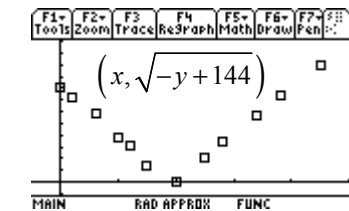
Applying the inverse of the apparent tool kit function to the y -values may produce linear data. In this case, if the data set (X, \sqrt{Y}) is linear, then that linear relationship is written as $\sqrt{Y} = m \cdot X + b$. A model for the

original data is found by solving for Y to get $Y = (m \cdot X + b)^2$. Members

of this family of quadratics have all positive y -values, and have their vertices on the x -axis. So before we reexpress, we need to move the data into this position – which requires that we know the y -value of the vertex. The transformation $Y = -y + y_{vertex}$ produces data that looks like this:



Taking the square-root of these new y -values produces this scatterplot: which is a problem... So: we delete the "left" half of the points to get linear data. Linear regression gives $m = 4.003$ and $b = 7.93$. (Look at residuals! It's fun to see how they look if you picked a "bad" vertex.) Solve to get a model for the original data:



$$\sqrt{-y + 144} = mx + b$$

$$y = 144 - (mx + b)^2$$

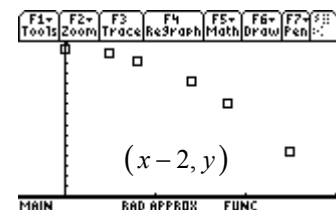
$$y \approx -16.02 \cdot x^2 + 63.49x + 81.12$$

Reexpression technique #2: "doing the x's"

Applying the apparent tool kit function to the x -values may produce linear data. In this case, if the data set (X^2, Y) is linear, then we get to write the equation $Y = m \cdot X^2 + b$ which is

also (the kids love this: no algebra needed!) the model for the original data. This family of quadratics has vertices on the y -axis. So our first task is to move the data into this position – which again requires that we know the vertex. The transformation $X = x - x_{vertex}$ produces this scatterplot:

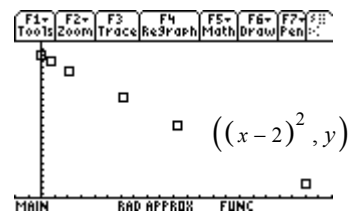
[Note: I only kept half of the parabola here, but later realized that I can keep all points. Why? Because squaring the x -values has no restrictions! Those points will just "flip over" the y -axis as a side-effect of the squaring.]



¹ I did screen shots from my TI-89, forgetting that most of you would have TI-83's... sorry!

Now we square the x 's, and see that $\left((x-2)^2, y\right)$ is indeed linear:

Linear regression gives $m = -16$ and $b = 143.2$
 so the model for the original data is $y = -16(x-2)^2 + 143.2$ or
 $y \approx -16x^2 + 64x + 79.2$.

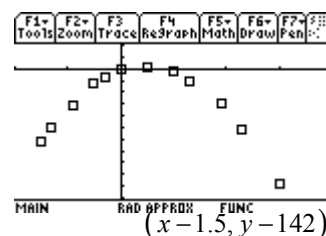


Reexpression technique #3

So what if you don't know the coordinates of the vertex? Consider the reexpression $\left(X, \frac{Y}{X}\right)$. If this data set

is linear, then we get to write the equation $\frac{Y}{X} = m \cdot X + b$ which leads to this model for the original data:

$Y = X(m \cdot X + b)$. Members of this family of quadratics are very loosely related: they all have two real roots, one of which is the origin. So for the reexpression to work, we only need to move the data so that any one point is, in fact, the origin. This is a very precise move – no guessing is needed! (But which point should you pick?) I picked the point $(1.5, 142)$ and made a new data set where $X = x - 1.5$ and $Y = y - 142$, producing this scatterplot:



Now graphing X versus $\frac{Y}{X}$ results in linear data:

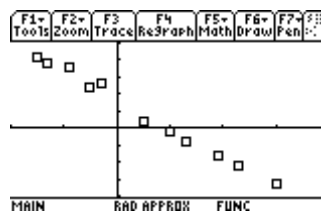
(What data point did I have to remove before doing this reexpression?)

Linear regression gives $m = -17.2$ and $b = 16.71$
 The linear equation is then solved for y :

$$\frac{y-142}{x-1.5} = m \cdot (x-1.5) + b$$

$$y = (m \cdot (x-1.5) + b)(x-1.5) + 142$$

$$y \approx -17.2x^2 + 68.31x + 78.23$$



$$\left(x-1.5, \frac{y-142}{x-1.5}\right)$$

There's a 4th way: Use Quad Regression on your TI to get $y = -16.12x^2 + 64.29x + 79.88$. This method is less valuable than reexpression if you want error bounds along with your model.

Summary

A comparison of all four models show that they are remarkably similar – because the data was remarkably nice! The reexpression $\left(X, \frac{Y}{X}\right)$ has the advantage of requiring no knowledge of the vertex, which is reflected in the fact that the vertex it creates is $(1.986, 146.1)$, not $(2, 144)$.

Follow-up

Give the students quadratic data where the vertex is not so obvious and have them try all four methods.