DERIVING A NON-LINEAR MULTIVARIABLE MODEL FOR STACKING NESTED CUBES VIA SPREADSHEET SIMULATIONS

Scott A. Sinex
Department of Physical Sciences and Engineering
Prince George's Community College
Largo, MD 20774
ssinex@pgcc.edu

Abstract

Spreadsheets are a powerful tool for engaging students to mathematically model data, at least use simulations to explore multivariable systems, and maybe construct them. Ideally instructors want to get students through the data > model > simulation path. To enhance the scientific aspects, experimental data collected by student measurements can be part of the process as well as considering measurement error. A variety of simple linear models such as stacking uniform sized objects and measuring the stack height allows students to hone their linear regression skills and build linear models. Here an activity is presented to introduce students to a non-linear system by measuring the height of a tower constructed from nested cubes (cubes that fit inside one another). The activity has students collect data, place it into a just-add-data spreadsheet and explore a linear model and goodness of fit using r-squared. They then examine the data to decide if the linear model is correct. They discover that a quadratic model is a better fit and see that curvature is revealed by the use of residuals. Using a pre-built spreadsheet simulation of the model, students derive a multivariable quadratic equation to describe the model. They further explore the dangers of extrapolation and measurement error.

Introduction

Mathematical modeling and the use of simulations are important parts of the *Common Core Standards in Mathematics* in K-12, and at the college level, have been endorsed by AMATYC and MAA. Here the use of an off-the-shelf piece of software produces an engaging pedagogy that brings algebraic and scientific thinking along with the use of hands-on manipulatives plus measurement and its error to the classroom. Spreadsheets and their computational power are an ideal tool for data > model > simulation for beginning students in both mathematics and the sciences. Mathematical modeling and simulations are tools to create an engaging pedagogy and deeper learning (Honey and Hilton, 2011).

A variety of simple experiments that develop linear mathematical modeling have been developed and summarized by Sinex (2013). These experiments get students into collecting data, dealing with measurement error, and handling linear regression analysis with the use of r-squared to judge the goodness of fit. Computational spreadsheet simulations are used to enrich the modeling experience. All of this requires the use of mathematics at the college algebra level and slowly enhances student's computational skills as they model data and use simulations.

This paper shows how via experimental height measurements and entering the data into a just-add-data spreadsheet students can produce a linear mathematical model of the height of a tower built by stacking cubes that are capable of nesting (fitting inside one another) and judge its goodness of fit. The objectives of this activity and accompanying interactive Excel spreadsheet or Excelet are to:

- examine a linear model with its goodness-of-fit judged by r^2 and reflect on the measurements and their possible errors;
- discover non-linear behavior (standard form of quadratic equation) and how to detect curvature of data by the use of residuals ($y_{measured} - y_{calculated}$);
- derive a model by numerical experimentation with multivariable parameters related to the behavior of nesting the cubes (make it more scientific from the quadratic parameters "a" and "b") via simulation;
- ascertain the dangers of extrapolation of the quadratic model, and;
- simulate possible measurement errors and how they influence the model including the masking of a small amount of curvature by random error.

This is a nice transition into non-linear models and the use of residuals to detect curvature in the data. Students revise the mathematical model to be multivariable from doing numerical experimentation, which addresses a number of practices (developing and using models, analyzing and interpreting data, using mathematics and computational thinking) in the *Next Generation Science Standards*.

Experimental Investigation

Using a set of nested cubes, students build a tower, starting with the largest cube at the base, and stacking the cubes. Height measurements to the nearest 0.1 cm (1 mm) should be made as each cube is added. Use of meter sticks is recommended since towers exceed 30 cm and the zero mark is at the end of the meter stick. An example with four cubes stacked is illustrated in Figure 1. After the data is recorded, it is entered into a just-add-data interactive Excel spreadsheet or Excelet, which generates a graph of the data, a linear mathematical model, and returns a value of r-squared. A screenshot of the "nested cubes" tab is shown in Figure 2. It includes a set of data measured by the author that yields a model with a fairly high value of r-squared and a large y-intercept. Students should examine the graph and decide if the linear model is appropriate plus reflect on the actual measurement and

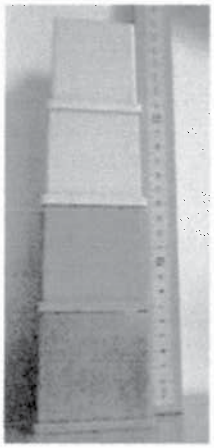

Figure 1: Tower

possible errors such as a tilted ruler (not vertical) during measurement. Eventually students are asked to change the plot to connect-the-dots with the data set; this makes it very apparent that the data is curved (not shown).

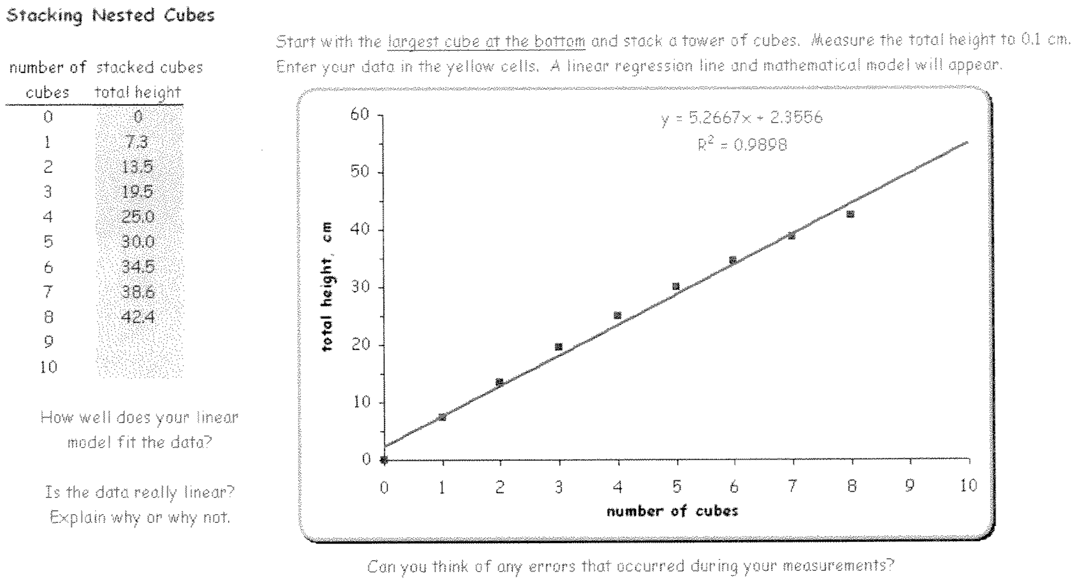

Figure 2: Nested Cubes Tab with Linear Regression

Students then fit the data with a quadratic regression and since the model should go through zero, the set intercept equal to zero (c = 0) model is used. A much better fit is obtained as shown below in Figure 3. For a review of a forced through zero regression and how to deal with it, see Eisenhauer (2003).

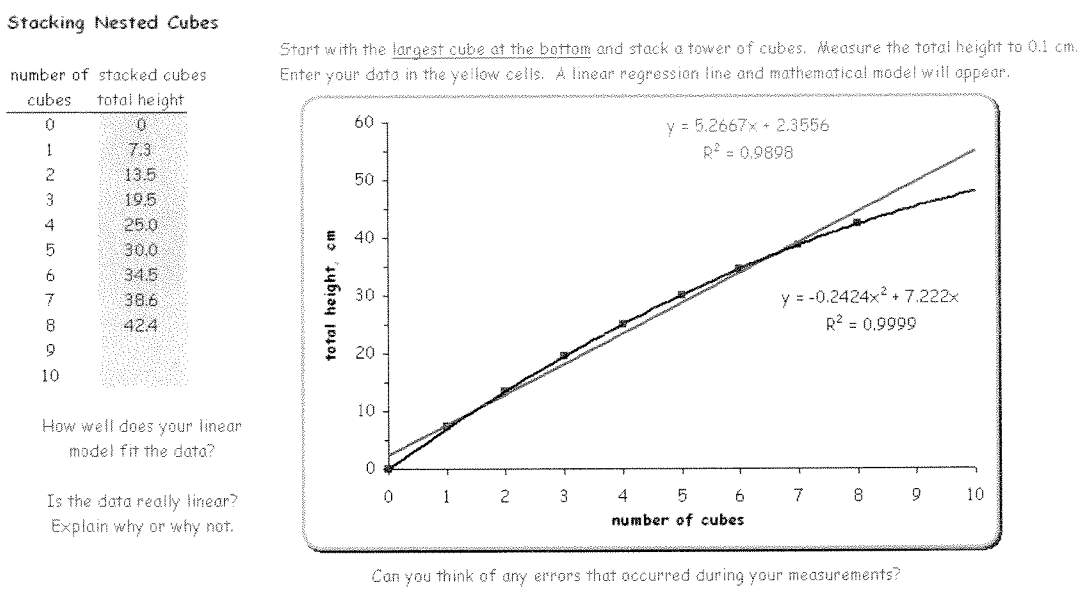

Figure 3: Nested Cubes Tab with Quadratic Regression (c = 0)

Now at this point, we want students to explore this quadratic model using a simulation in Excel.

<u>Simulation of Stacking Nested Cubes Model</u>

To build on some science process with the mathematical model of the data, the "simulation" tab given in Figure 4 will include a simple all same height cubes data and its model (in red) and both the linear and quadratic models for the nested cubes (in blue). Two new variables have been introduced in the simulation, the "initial edge length" of the first or largest cube is given in the yellow cell and the "decrease the edge of the next cube by" (space needed to nest) by the scroll bar. Also given below the graph of the data and models is the residuals graph for the blue linear model of the data. The residual is the actual y-value minus the calculated y-value. Students should notice the nice curve shown for the residuals indicating a non-linear fit.

Stacking Nested Cubes - The Simulation

number of cubes	nested cubes edge, cm	nested cubes total height	all same height
0	0	0	0
1	10	10	10
2	9.5	19.5	20
3	9	28.5	30
4	8.5	37	40
5	8	45	50
6	7.5	52.5	60
7	7	59.5	70
8	6.5	66	80
9	6	72	90
10	5.5	77.5	100
	↑ size of each nested cube		↑ constant size cube

The yellow cell above for the initial cube edge (5-25) can be adjusted.

Residuals show curvature of blue linear model

Sinex 2014

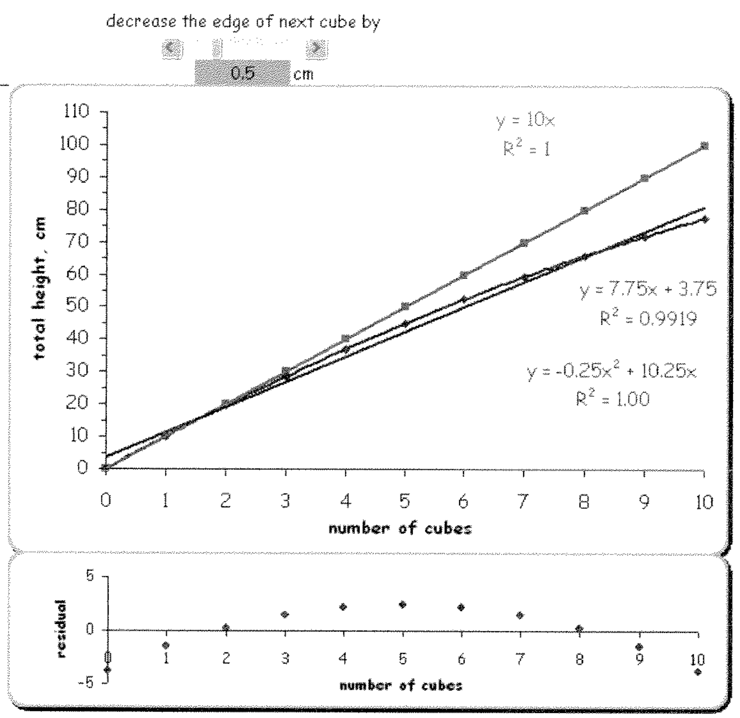

Figure 4: Simulation Tab

The goal here is to take the quadratic regression equation as given by Excel and first convert it to the variables measured in the experiment.

$y = ax^2 + bx + c$ and $c = 0$ → $y = -0.25x^2 + 10.25x$ → $H = -0.25n^2 + 10.25n$

where H = height and n = number of cubes

By using the simple all same height cubes model, students should be able to figure out that the slope is the edge length of the cubes. This helps them realize that a parameter such as the edge length of a cube (e), H = en, can be a new variable introduced into an equation.

Now using the simulation students can explore the influence of the "initial edge length," e, and the "decrease the edge of the next cube by," d, to see how these variables can be incorporated into the model equation:

$$H = (-d/2)n^2 + (e+d/2)n$$

Hence an equation is found, totally given in scientific variables, that discloses the multivariable nature of the mathematical model. We also have students check the set intercept equal to zero model by doing a standard quadratic regression to see that the "c" is very small and can be neglected. This is suggested by Eisenhauer (2003).

Once students derive the model, they can set up the computations to examine what happens on extrapolation. They find that the model hits a maximum and then decreases with larger number of cubes. Students also examine the residuals for the quadratic model with c = 0 to see the more random distribution and much smaller magnitude. This is done using Solver in the pre-set mode to minimize the sum of the squared errors (residuals) or SSE.

Investigating Simulated Errors

With the "simulated errors" tab, we can investigate ruler error, which is the position of the zero mark on the ruler, and random measurement error. The key to investigating an error is to investigate them one at a time and set the ones not being investigated to zero error. So first up to experiment with is the ruler error. How does adding ruler error influence the model? To discover this we will examine three cases with the first one being the most straightforward to detect the error. The screenshot in Figure 5 uses a typical quadratic regression and does not include the origin (0, 0) in the data set. The regression results for e = 10 cm and d = 0.4 cm is given below as:

$$H = -0.20n^2 + 10.20n - 1.00$$

Why not consider the origin? The origin is assumed and not measured, so it would not have any error associated with it. So let's see how it influences the regression and, hence, the detection of the error. Here are the other possible two models.

For a standard quadratic regression with the origin as a data point (but no error in this point):

$$H = -0.18n^2 + 9.98n - 0.42 \qquad r^2 = 1.000$$

where the c = -1.00cm and this is the distance the zero mark is from the end of the ruler, which decreases all height measurements by 1.00 cm. The constant cube size shows this easily. The model above is simply translated down by 1.00cm on the graph. This is a negative constant systematic error. For more on the type of errors, see Sinex (2005a) for a simple introduction and Sinex, Chambers, and Halpern (2012) for an experimental approach.

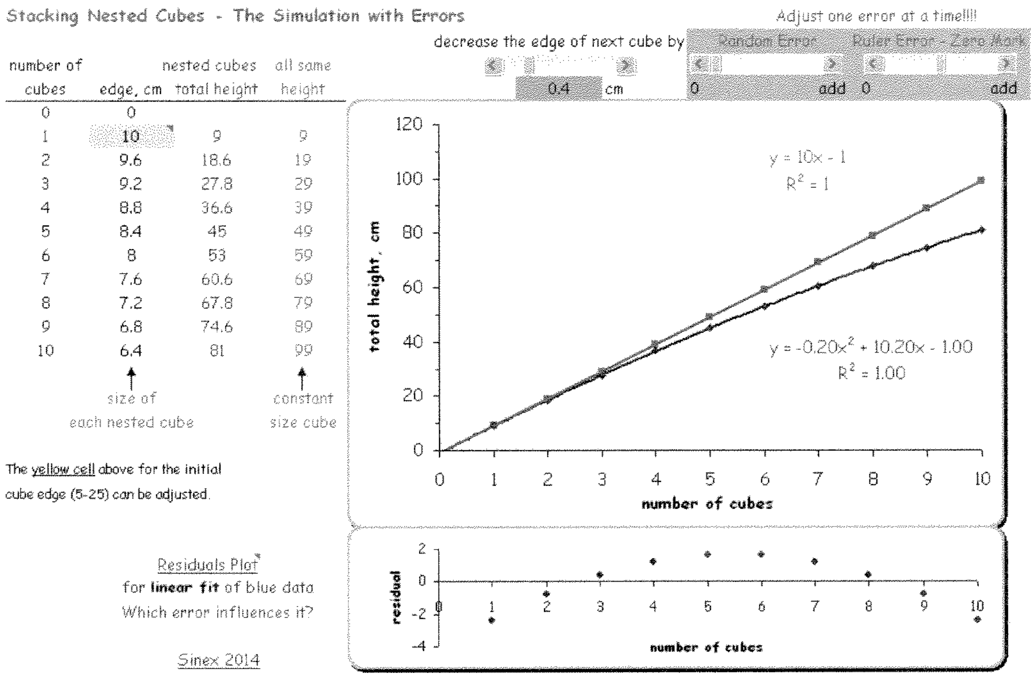

Figure 5: Simulated Errors Tab – Ruler Error (Systematic error)

This model translates down a smaller extent plus it rotates downward as well.

For a quadratic regression with the set intercept equal to zero:

$$H = -0.17n^2 + 9.82n \quad r^2 = 1.000$$

This model cannot translate, since it is forced through zero (c = 0), so it rotates downward even more than the second model. These two situations make detecting the negative constant systematic error much more difficult. This error does not influence the r^2 value or the residuals.

For more about ruler errors and the adventures of the Dummass Ruler Company, see the Rulers and Measurement Errors Excelet at http://academic.pgcc.edu/~ssinex/excelets/Ruler_error.xls.

Now what about random error? Random error introduces scatter into the data points and essentially introduces randomness into the values of the regression parameters (a, b, and c). Random error reduces the value of r-squared and adds scatter to the residues plot. This is illustrated on the screenshot in Figure 6. Students should also notice (this may take multiple trials to see in Excel, so have students play) that for the linear model (not shown on upper graph), the residuals plot (bottom graph for the linear model) looks random and hence you might miss detecting the curvature in the data as it is masked by random error. For more on dealing with scatter, r-squared, and residuals, see Sinex (2005b).

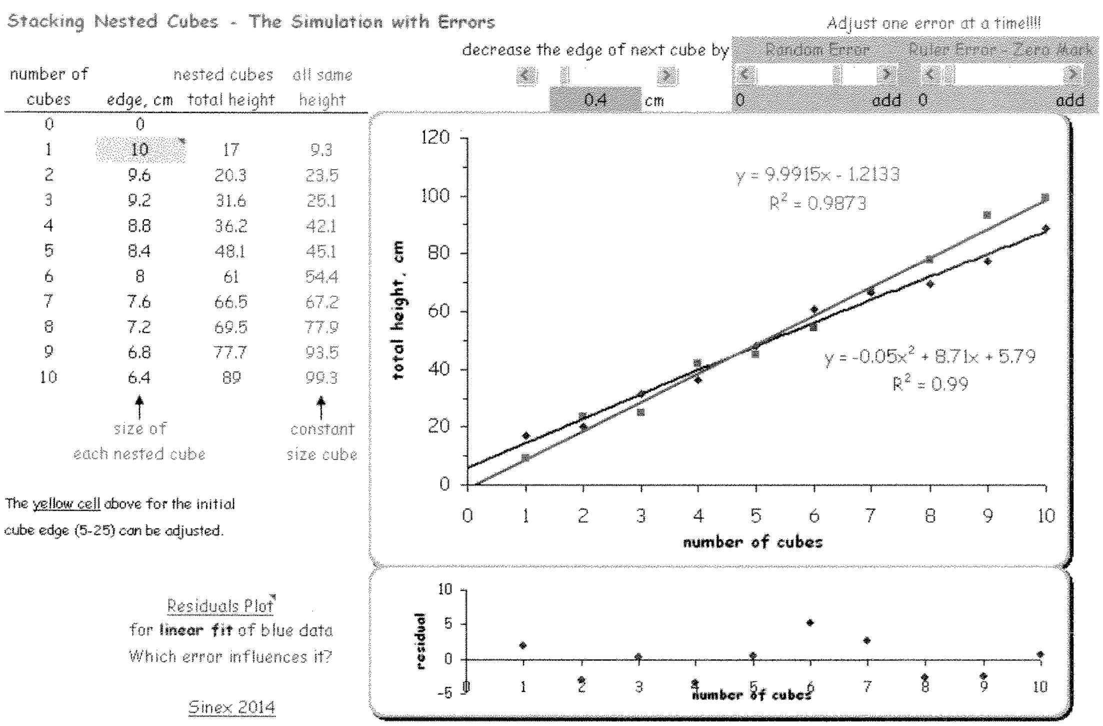

Figure 6: Simulated Error Tab – Random error

The complete evolution of the model is summarized in Table 1 on the next page.

Mechanics of the Simulation and Errors

The simulation of the "all same height" cubes is shown in Figure 7. In cell D11 = A11*$B$6 (shown on formula bar at top of figure) or H = n*e. This is shown by using trace precedents on the worksheet for two calculations.

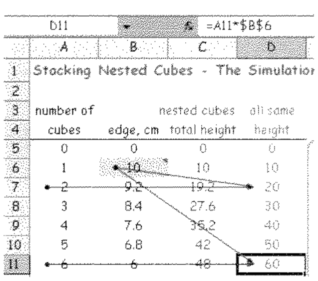

Figure 7: All same height

To do the "edge" and "nested cubes total height," the formulas use the simple arithmetic of the edge and not the model regression equation as illustrated on the screenshots below (edge- Figure 8; nested cubes total height- Figure 9).

Table 1: Evolution of the Mathematical Model for Stacking Nested Cubes Tower

Type of Model	Equation	Variables	Goodness of Fit	Comments
Linear	$y = mx + b$ $H = mn + b$	m = slope b = y-intercept <u>Experimental</u>: H = height (cm) n = number of cubes	r-squared is high <u>Better method</u>: residuals plot shows a very curved pattern	y-intercept should be near zero and it's not! **Is this model correct?**
Standard form quadratic	$y = ax^2 + bx + c$ $H = an^2 + bn + c$	a, b, & c are parameters to fit quadratic	r-squared is higher; residuals plot is now random	small value of "c" due to measurement error
Standard form quadratic with c = 0 (Set intercept to zero in Excel)	$y = ax^2 + bx$ $H = an^2 + bn$ in terms of more model variables: $H = (-d/2)n^2 + (e+d/2)n$	a and b are parameters to fit quadratic <u>More Experimental</u>: e = initial cube edge length d = distance to decrease edge to nest	if a meter stick is used to measure (zero at end of ruler), results should be very similar to above	y-intercept should be zero, so c = 0 with more experimental variables, equation shows multivariable nature of model
Standard form quadratic with systematic error  Three cases to examine	$y = ax^2 + bx + c$ without origin in data $y = ax^2 + bx + c$ with origin in data $y = ax^2 + bx$ with fixed c = 0	as above to explore quadratic behavior **How does error influence the model?** "c" should equate to the distance the zero mark is from the end of the ruler ($c < 0$)	r-squared and residuals plot are NOT influenced by ruler error for all three cases! – this is a negative constant systematic error	w/o origin in data, curve translates down by "c" w/ origin, smaller drop in c, a & b decrease as curve rotates downward w/ fixed c = 0, a & b decrease even more w/ greater downward rotation
Standard form quadratic with random error	$y = ax^2 + bx + c$ add random noise to "y" variable (height)	causes random variation of a, b, and c	r-squared is lowered; residuals plot is more scattered	can mask small amount of curvature in residuals for linear models!

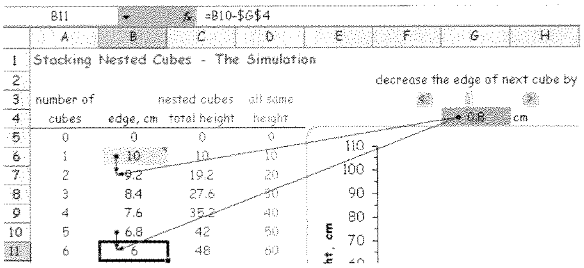

Figure 8: Edge

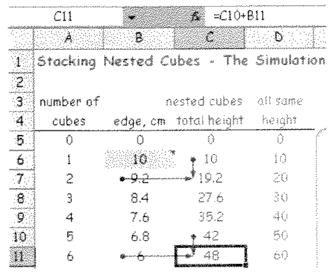

Figure 9: Nested Cube Height

To view all the computational formulas, go to the spreadsheet and do [Ctrl] [`] and the formulas in all cells appear then repeat key stokes to return to numerical values.

For the "simulated errors" tab to calculate the random and negative constant systematic errors:

in cell C6 =B6-$L$3/10+$J$3/100*RANDBETWEEN(-10,10)

where -$L$3/10 is the negative constant systematic error (only appears in cell C6 due to the nature of the calculation) and $J$3/100*RANDBETWEEN(-10,10) adds the random noise in the "nested cubes total height" column. View the "all same height" column that illustrates a more typical set-up where the error formula appears in each row. The $L$3 and $J$3 values adjust the size of the error and usually require some experimentation to decide on the magnitude of the error. The RANDBETWEEN function generates a random number between -10 and 10 in this case. The mechanics of how-to add all the various errors can be found in Sinex (2013).

Some Final Thoughts

A hands-on minds-on activity using an engaging pedagogy with an interactive Excel spreadsheet or Excelet is presented for a quadratic model. Students must collect data by making careful measurements, consider a linear model, and then discover that a quadratic mode is a better choice from using residuals as a method to judge goodness of fit. Students discover by exploration of a simulation a more scientific model that is now multivariable. Students also consider errors and the dangers of extrapolation.

For those who want to produce their own interactive Excel spreadsheets, see the Developer's Guide to Excelets at http://academic.pgcc.edu/~ssinex/excelets. Over a hundred pre-built spreadsheets are available free to download plus activities, instructions, and links to other spreadsheet sites. Look under the graphs for notes on mechanics and some calculations.

The activity for this paper including the Excelet can be downloaded from http://academic.pgcc.edu/~ssinex/excelets/nested_cubes_act.pdf.

Acknowledgements
My sincere thanks to Barbara Gage, Professor Emerita at Prince George's Community College, for providing her invaluable input to this work.

References

Common Core Standards for Mathematics, http://www.corestandards.org/Math.

Eisenhauer, J.G. (2003) Regression through the Origin, **Teaching Statistics 25** (3) 76-80.

Honey, M.A. and Hilton, M.L., Eds. (2011) *Learning Science through Computer Games and Simulations*, National Research Council, National Academy Press.

Next Generation Science Standards, http://www.nextgenscience.org/.

Sinex, S.A. (2005a) Investigating Types of Errors, **Spreadsheets in Education 2** (1) Article 7.

Sinex, S.A. (2005b) Exploring the Goodness of Fit in Linear Models, **Journal of Online Mathematics and its Application 5.**

Sinex, S.A. (2013) Investigating and Visualizing Measurement Error for Novice "STEM" Learners in *Proceedings of the Spring 2013 Mid-Atlantic Section Conference of the American Society of Engineering Education*, pp. 222-234.

Sinex, S.A., Chambers, T.L., and Halpern, J.B. (2012) Mass, Measurement, Materials, and More Mathematical Modeling: The Nuts and Bolts of Let's Make an Error, **Spreadsheets in Education 6** (1), Article 5.