January 2006

Investigating Types of Errors

Scott A. Sinex
sinexsa@pgcc.edu

Recommended Citation

Investigating Types of Errors

Abstract

An interactive Excel spreadsheet investigation of random and systematic errors including the effect of curvature on linear models is presented for introductory-level science classes. Students are engaged in higher-order thinking and science process skills as they navigate through the activity.

Keywords
random error, systematic error, discovery learning, interactive Excel spreadsheet

Investigating Types of Errors

Scott A Sinex

Department of Physical Sciences and Engineering
Prince George's Community College
ssinex@pgcc.edu

Abstract

An interactive Excel spreadsheet investigation of random and systematic errors including the effect of curvature on linear models is presented for introductory-level science classes. Students are engaged in higher-order thinking and science process skills as they navigate through the activity.

Keywords: random error, systematic error, discovery learning, interactive Excel spreadsheet.

1.   Introduction

Getting students in introductory-level science courses to understand how to deal with error is a challenge to most instructors; however, error should always be considered when collecting and analyzing data. Error in experimental measurement can be categorized as either random or systematic. Random errors will fall both above and below a true value (tv ± e) or cause scatter in results. As more measurements are made the average value approaches the true value. Systematic errors on the other hand generate either a positive or negative bias in results. Results are consistently above or below the true value. Over a range of measurement values, systematic errors may be constant (tv + e or tv – e) or proportional (e·tv where e > 1 or e < 1). See Kirkup [3] for a thorough discussion of handling errors. Here we simulate a variety of errors providing students with an avenue to explore how the errors influence the behavior of data on a graph. With data sets available from a number of sources and collection of data using graphing calculators and computers readily accomplished, it is easier to take a mathematical modeling approach in mathematics [2, 4] and science classrooms [4, 7]. Judging goodness of fit using the coefficient of determination, r^2, has become common practice [8]. We can now ask "How do various errors influence a series of measurements?" and "How is r^2 influenced by errors?".

Students will use the accompanying interactive Excel spreadsheet and student activity to explore the effects of random and systematic errors on a linear model. Both proportional and constant systematic errors are included. The effects of curvature in general and at the high end of the x-variable are also explored. We are going to use the spreadsheet as a discovery learning tool to investigate the variety of errors by comparing the errant data with the y = x line. Discovery learning is in line with national reform efforts in science, as well as mathematics, education in the United States [2, 4, 5, 7]. From an experimental point of view, using two measurement techniques over a range of values works or two different analysts doing the same measurement over a range provides a method to discover error and

for comparison of error behavior. A number of experimental possibilities are given in Table 1 below.

Table 1: Some Possible Comparison Measurements

Measurement	Technique 1	Technique 2
Dissolved oxygen	Winkler method	Membrane probe
Mass of objects	Electronic balance	Spring scale
Distance - time	Stop watch	Photogates
Height	Meter stick	Motion detector
Spherical object diameter	Calipers	Ruler

Calibrations of instruments in a variety of fields are other possibilities; however, a new "errorless" comparison is needed in place of the y = x line.

2.   Student Exploration of the Interactive Spreadsheet

Students are guided through the use of the spreadsheet by the mode of questioning (See the Student Activity where a comparison of masses on different balances is given). We are assuming that students are familiar with handling, graphing, and analyzing data in Excel. We provide support documentation for students [9]. How does the data on the graph respond to the variety of possible errors? One explores each error individually to discover the effect of that error. Adjustable sliders or scroll bars are used to change the variable errors. A combination of errors, a more realistic scenario, can then be investigated. We also include the effect of the data being non-linear. This requires the use of residuals, since the r^2 is not going the reveal curvature. A plot of the residuals is given on the second worksheet of the spreadsheet. Investigating the effects of the individual errors is summarized by students, who fill in Table 2 which is blank on the student activity. The coefficient of determination, as given by r^2, is the fraction of the *y*-variable that is explained by the variation of the *x*-variable; it ranges from 0 to 1. Here $r^2 = R^2$ as used in Excel, where R is the Pearson product moment correction coefficient (search RSQ in Excel help).

        The nature of random and systematic errors can be discovered via exploration of the spreadsheet. After the individual errors are investigated, combinations are examined. Since systematic errors can usually be corrected, we explore this possibility. Asking students to examine the combinations as in Figure 1 and determine what the data would look like after any correction applied is a good assessment tool to assure understanding.

116

Table 2: Effects of the Errors

Error	Graphical response	r^2	Residuals
Random	Causes scatter in data points	Decreases value	Random distribution of points
Proportional systematic	Changes slope of graph	No effect	No effect
Constant systematic	Changes y-intercept of graph	No effect	No effect
Minimal curvature in data	Not easily visible at small amounts	Very minimal decrease at small amounts	Pattern appears in distribution of points
"Bad cal" or curvature at larger x values	Not easily visible at small amounts	Very minimal decrease at small amounts	Pattern appears in distribution of points
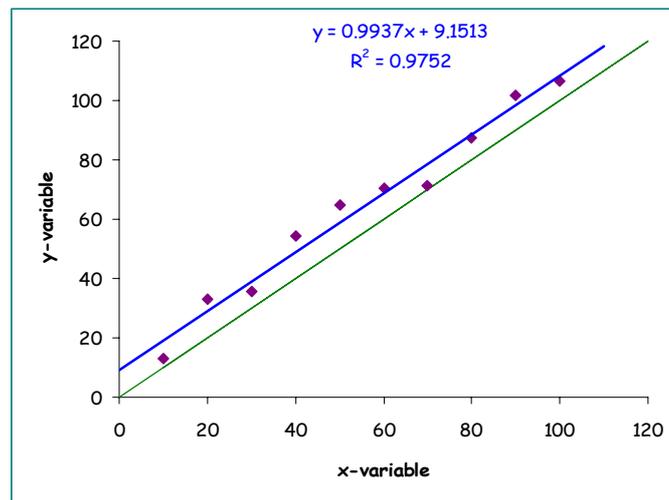
Figure 1: A combination of random and constant systematic errors

One can also consider measuring devices (Figure 2) and explore the type of errors produced by them. How would they influence a variety of length measurements? Both rulers cause a constant systematic error (A is below true value, B is above). Adding random error to a plot with a small amount of curvature is interesting for students to examine. Recognizing curvature can be difficult since random error can disguise the curvature. Further investigation of curvature can occur with non-linear models as students advance in their modeling skills.
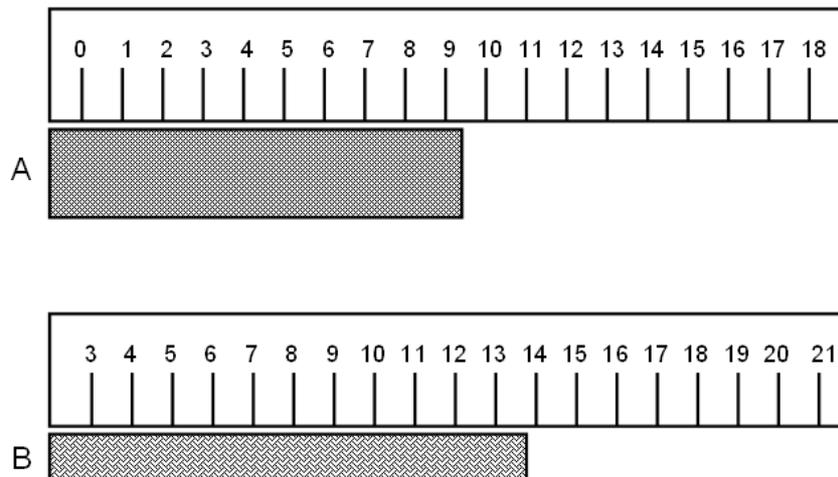
Figure 2: Errant Rulers

3. Mechanics of the Errors

The errors are controlled by including an additional variable for each type as shown in the equation below:

$$y = ax + bRANDBETWEEN(-10,10) + c + dx^2/5000$$

where a is the proportional error, b adds random noise using the RANDBETWEEN[1] function in Excel, c is the constant error, and d adds a small amount of curvature to the data. All the error variables are linked to adjustable scroll bars to control their magnitudes. For the Bad Cal error an additional term appears for x = 70 and greater.

$$y = ax + bRANDBETWEEN(-10,10) + c + dx^2/5000 - ex^2/22000$$

Here the amount of curvature increases by lowering 22000 at x = 70 to 13000 at x = 80, 10000 at x = 90, and 8000 at x = 100 (See cells B12-B15 on the errors worksheet). As the value of e is increased the high end of the plotted data starts to bend downward. This is a case of the linear model breaking down and is common in calibration curves in spectrophotometry.

To aid in the interpretation of residuals we have added conditional formatting, a highly underused feature [1], to this column of data. This allows the pattern or lack there of to be seen by color coding of the cell background in the numerical data in addition to using a graphical approach. This is accomplished by going to Format on the menu bar and selecting Conditional Formatting… and the pop-up window shown below in Figure 3 appears. The use of conditional formatting enhances the random nature of the residuals when exploring the random error as well. This approach works to strengthen the "rule of four," where mathematics educators are trying to emphasize and relate graphical, numerical, symbolic, and verbal descriptions [2, 5].

---

[1] The use of the RANDBETWEEN function requires the Analysis ToolPak to be installed in Excel (search for Analysis ToolPak in Help for loading instructions).
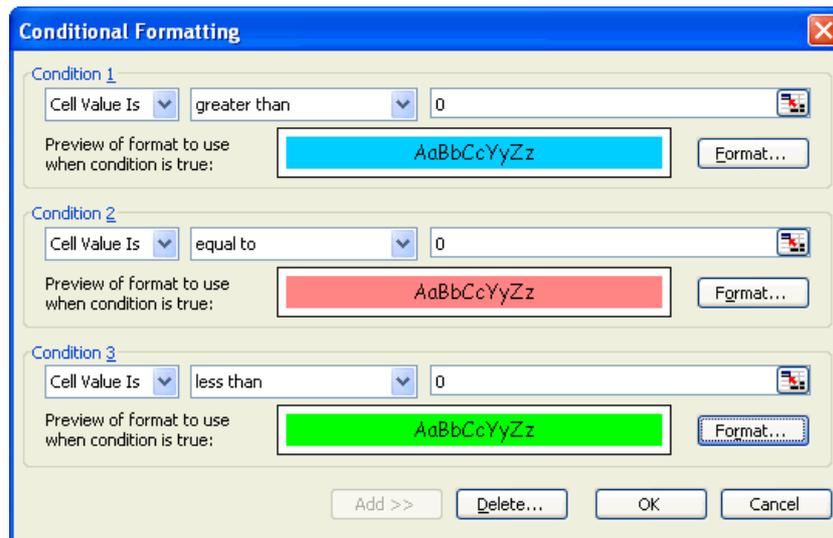
Figure 3: Conditional Formatting pop-up window

4.   Some Final Thoughts

Exploring the behavior of random and systematic errors can be done quickly and easily with this spreadsheet. Using simulations and "what if" scenarios, novice learners can start developing the mind-set to consider error in data and discover how errors influence data [6]. The spreadsheet can be used individually by students or projected for use in a classroom. The graphical, numerical, symbolic, and verbal descriptions are all integrated into this activity [2, 5]. The numerical experimentation is, in itself, science process. We use this in the beginning of second semester general chemistry as part of series of experiments to get students in the habit of thinking about and handling errors. We have also developed a hands-on activity to enable students to get a physical handle on error using a simple, understandable procedure involving sandwich cookies [10].

References

[1]   Abramovich, S. and Sugden, S.J. (2004). Spreadsheet Conditional Formatting: An Untapped Resource for Mathematics Education. *Spreadsheets in Education 1* (2): 85-105.

[2]   AMATYC (2004). *AMATYC Standards 2006 Beyond Crossroads*. American Mathematical Association of Two Year Colleges (online).
http://www.amatyc.org/Crossroads/CRRV6/BC_V6_home.htm
accessed 3 October 2005.

[3]   Kirkup, L. (2002). Data Analysis with Excel: An Introduction for Physical Scientists. Cambridge, UK: Cambridge University Press.

[4]   National Research Council (1996) *National Science Education Standards*. Washington, DC: National Academy Press.

[5]   NCTM (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

[6]   Sherman, T., and Kurshan, B. (2004). Teaching for Understanding. *Learning and Leading with Technology 32* (4): 6-11.

[7]   Siebert, E.D. and McIntosh, W.J. (2001). *College Pathways to the Science Education Standards*. Arlington, VA: National Science Teachers Association Press.

[8]   Sinex, S.A. (2005). Exploring the Goodness of Fit in Linear Models. *Journal of Online Mathematics and its Applications 5* (online).
http://www.joma.org/mathDL/4/?pa=content&sa=viewDocument&nodeId=559
accessed 3 October 2005.

[9]   Sinex, S.A. and Gage, B.A. (2001) Using Excel for Handling, Graphing, and Analyzing Scientific Data: A Resource for Science and Mathematics Students (online).
http://academic.pgcc.edu/psc/Excel_booklet.pdf accessed 3 October 2005.

[10] Sinex, S.A., Gage, B.A., and Beck, P.J. (2005). Exploring Measurement Error with Cookies: A Real and Virtual Approach via Interactive Excel. *AMATYC Review*, submitted.

120

Appendix 1: Student Activity

Investigating Types of Errors: Checks and Balances

This activity uses the interactive Excel file, *investigating types of errors.xls*, to explore a variety of types of errors and how they influence data. Let's consider a scientist checking a set of standard masses on four different laboratory balances. We will assume each standard mass is the true value with no error. For each balance below, plot the balance mass as a function of the standard mass and perform a linear regression on each set.

Standards (grams)	Balance 1	Balance 2	Balance 3	Balance 4
2.00	2.00	4.00	1.70	1.81
5.00	5.00	7.00	4.25	5.57
10.00	10.00	12.00	8.50	9.57
15.00	15.00	17.00	12.75	15.67
20.00	20.00	22.00	17.00	20.89
25.00	25.00	27.00	21.25	24.12

Record the regression equation and value of r^2 in the table below.

	Balance 1	Balance 2	Balance 3	Balance 4
Linear regression equation				
r^2				
Describe the error				

For two variables that are in total agreement (no error) the y = x line fits the data. This is the case for Balance 1 above. There is no error in the mass determination on Balance 1. For Balance 1 the equation is $mass_{balance} = mass_{standard}$. This is a common comparison method we are going to use to investigate the possible errors that could occur. With this reference in mind, describe the errors in Balances 2, 3, and 4. You might want to plot these and compare them to the y = x line.

Now go to the interactive Excel spreadsheet. The errors worksheet allows you to explore a variety of errors. Residuals are the difference, Δy, of y-measured minus y-calculated from the regression equation. The residuals worksheet responds to the errors.

Fill in the information about the various errors. The data should start out lying on the y = x line. Adjust one parameter at a time while holding all the others to no error. For each type of error, consider how the slope, y-intercept, and r^2 respond. For the residuals plot, look for patterns or lack of patterns in the data.

Error	How does the graph respond compared to the y = x line?	How does r^2 change if at all?	View the residuals graph and explain how the points are distributed.
Random			
Proportional systematic			
Constant systematic			
Minimal curvature in data			
"Bad cal" or curvature at larger x values			

What type of error occurred for the balances?

	Balance 2	Balance 3	Balance 4
Type of error			

Now in the real world combinations of errors occur. What are the errors exhibited on the graph below?
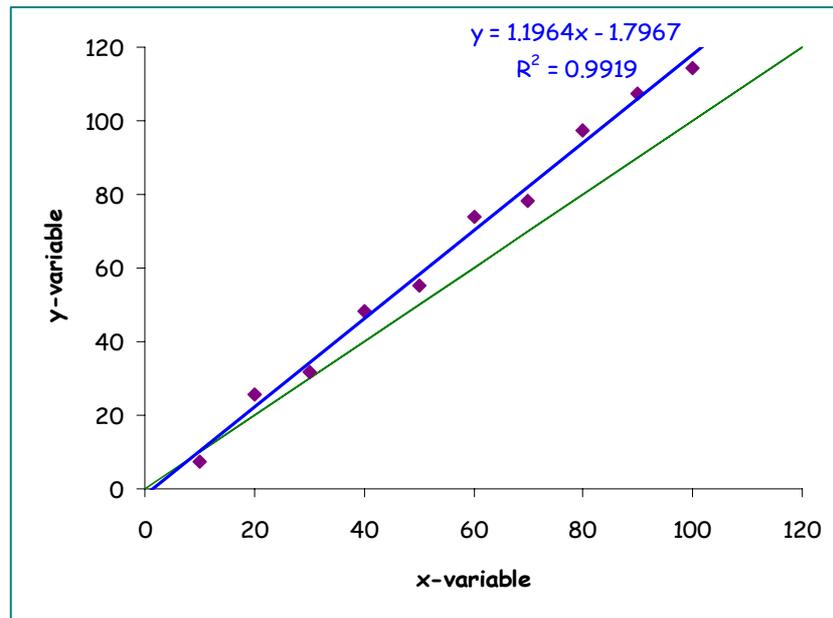
Systematic errors are usually correctable after investigation. They may involve a miscalibrated measuring device or instrument. How would the graph above change if the systematic error was corrected?

Now place a small amount of curvature into the spreadsheet with all the other error set at none. Observe the residuals graph. What happens to the residuals in this case?

Could you discover the curvature from the r^2 value?

Curvature shows up in the residuals as a pattern in their distribution. The r^2 value changes (decreases) very little for a small amount of curvature. A good linear fit should show a random pattern in the residuals. Increase the random error in this data with a small amount of curvature. What happens to the residuals?

Random error can disguise the non-linear character of a data set. Random counting error in long-lived radioactive material could disguise the exponential nature of radioactive decay.

124