Statistics for Success

Statistical analysis of student data is a lot easier than you think and more useful than you imagine.

By Robert Kadel

Subject: Assessment, data-driven decision making

Audience: Teachers, teacher educators, technology facilitators, staff developers

Grade Level: K–12 (Ages 5–18)

Technology: Spreadsheets

Standards: *NETS•T* IV (http://www.iste.org/standards/)

To her surprise, Ms. Logan had just conducted a statistical analysis of her 10th grade biology students' quiz scores. The results indicated that she needed to reinforce mitosis before the students took the high-school proficiency test in three weeks, as required by the state. "Oh! That's easy!" She exclaimed.

Teachers like Ms. Logan are everywhere. You may be a lot like her yourself—you have lots of data on your students: quiz scores, homework grades, attendance records, and so on—but aside from reporting such information to your school, you don't know what else to do with it.

But conduct a statistical analysis? Why in the world would you want to do something like that? Well, aside from making you the life of staff parties, you might be surprised at just how much you can help yourself and your students with some simple statistical procedures. Using statistics to look at student data can help you to prioritize your time and energy, give you a picture of student progress, and allow you to understand the relationships between, say, attendance and achievement or homework completion and achievement.

Statistics do not have to be the big, bad monster you may have thought of when faced with the course in college. The basic concepts are simple. Student quiz scores are data; the class average on a quiz is a statistic. Consider what you would learn if you had class averages on six quizzes covering six different topics. This information could tell you where your students are doing well and where they need remediation.

Furthermore, suppose you have two classes of students, where one class is using new software in addition to the textbook for exploring science concepts while the other class is using just the textbook. Calculating the quiz averages across six different topics for *each* class can help you discover areas where the software is more effective and where it is not.

That said, we can now turn to using Excel to take on just this type of analysis. You can use other spreadsheet programs; most work very similarly to Excel. Figure 1 shows a simple, sample grade book: 12 students' scores for six different quizzes in a high school biology class. Of course, the concept works for virtually any curricular area.

First, let's consider average quiz scores. Click to highlight the box (a.k.a. a "cell") at the bottom of Column B, the column of numbers labeled "Intro & Scientific Method." Then click the "Paste Function" button (it has *fx* on it). A dialog box opens that says "Paste Function." In that box, click "Statistical" on the left, and then click "AVERAGE" on the right. Then click "OK."

Because you clicked in the cell at the bottom of Column B, Excel automatically assumes that you want to find the average of the numbers above it. That's why it puts B2:B13 in the "Number 1" blank. Note that cell B1 has the heading, the name of the quiz.

You can change this either by typing new cell references, such as B2:B11, to get the average for just the first 10 students, or C2:C13 to get the average score for the quiz on Spontaneous Generation Theory. Finally, click "OK" calculate the average quiz score. (In this case, 84.16667.)

Now here's the really easy part. Move your mouse so that the cursor hovers over the lower-right corner of the cell that shows the average quiz score. The cursor will turn into a plus sign (Figure 2). This means that you should click-and-drag the cursor to the right until you've highlighted all the cells at the bottom of each column of quiz scores. When you release the mouse button, Excel will automatically fill in those cells with the average quiz scores for each topic.

If it looks a little busy to you because of all the decimal places, make sure all those cells are still selected, then go to the "Format" menu, select "Cells…," and then on the "Number" tab, select "Number." From there, you can change the number of decimal places to 1 or 2 and have a much more manageable view of the averages you just calculated.

So what areas need the most remediation before the students take their year-end science exam? With just six subjects listed here, you can probably eyeball the correct response. If you have many more, you may want to create a chart to help you get a graphical representation of the statistics.

Creating a chart in Excel is reasonably easy to do, but it takes some planning. Start by doing two things. First, click-and-drag your mouse across the cells in Row 1 of the spreadsheet where the titles of the six lessons appear (from "Intro & Scientific Method" to "Mitosis"). Then click the Copy button or go the "Edit" menu and select "Copy." Next, click on an empty cell anywhere on this spreadsheet (or on another spreadsheet). Then click the Paste button or go to the "Edit" menu and select "Paste."

You've copied your lesson titles, and now you need to copy the averages you just calculated. This procedure is almost the same (select the cells with the averages in them, click Copy), but you need to go to the "Edit" menu and choose "Paste Special…" when you get the dialog box shown in Figure 3. Select "Values" and then click "OK."

Now that you have those groups of cells next to each other, you can create a chart quite easily. First, click-and-drag your mouse from "Intro & Scientific Method" to "55" to select all the cells. Click the Chart Wizard button or go to the "Insert" menu and select "Chart…" The first dialog box in the Chart Wizard asks you to select a "Chart sub-type." I suggest leaving it at the default setting for now, which is a two-dimensional bar chart. Then click "Next." The second dialog box is easy, because you've already set up your data for it. This is where you would select multiple rows and columns of cells to create your chart; but because you've already put your lesson titles together with their average scores and then highlighted that group of cells, Excel defaults to that selection for the chart. So you can click "Next" here too, and move on.

The third dialog box allows you to change a number of chart options. On the "Titles" tab, the blanks for chart title, category axis, and value axis are blank by default. Type in some information here and you'll see the preview of your chart update to include a title. You may wish to explore the options on some of the other tabs; for example, the Data Labels tab will allow you to put the number you are graphing (average quiz score) at the top of each bar.

When you are finished with the Chart Options box, click "Next." The final dialog box asks if you would like to have Excel create the chart on a new sheet or insert it into the existing worksheet. I suggest you choose "As new sheet:" for the time being. (You can experiment with the layout of your chart when embedded into the current worksheet at a later time.)

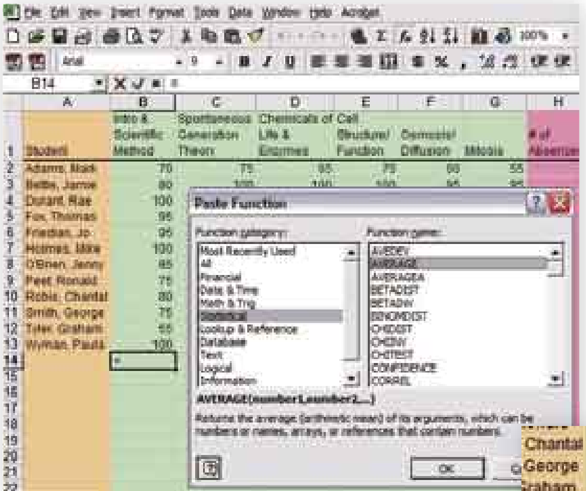

Figure 1. A typical gradebook spreadsheet in Excel. The dialog box shows that Excel is ready to paste the average scores from the first column of quiz results.

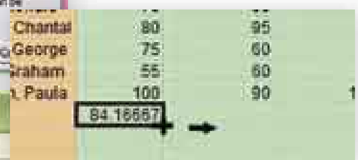

Figure 2. By dragging on the average box, you can have Excel perform the same function on each column of scores.

Figure 3. When creating a graph, it is important to use the "paste special" function in order to preserve the values of your data.

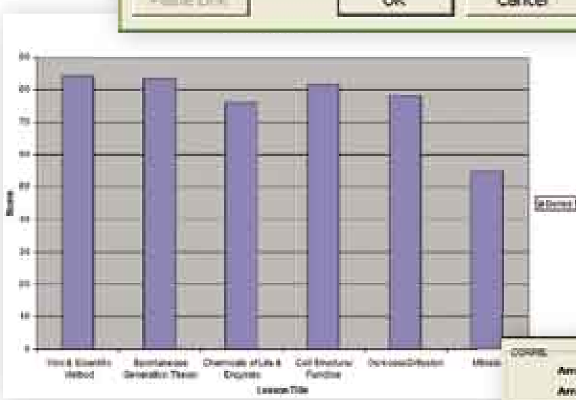

Figure 4. A simple chart created in Excel can make your statistics stand out.

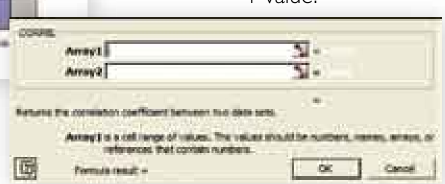

Figure 5. Exploring correlations is easy. Simply place the variables in the array boxes and discover the r value.

Figure 6. The gradebook spreadsheet with one variable (Array1) selected (quiz results for Intro & Scientific Method). To correlate these scores to # of Absences, you'd highlight the data in that column for Array2.

Click "Finish" and you'll have a chart that looks like the one in Figure 4.

Looking at the chart, one can conclude that students need some extra attention in learning mitosis. They may also need some help with "Chemicals of Life & Enzymes" and "Osmosis/Diffusion," but clearly mitosis is the priority here.

Correlations

Let's cover one more area where Excel can be helpful, and that's in understanding the relationship between two different kinds of student data. Statisticians call this correlation, and it's a very handy bit of information. First, I'll cover the statistical basis, and then I'll show you how to calculate a correlation in Excel.

The concept of correlation comes from taking two things you can measure, such as quiz scores and attendance, and seeing if there is a relationship between them. Lots of people believe that there is a relationship between attendance and grades, and usually it's that as absences increase, grades decrease. But that's not always the case, and if you want to know for sure, you need to look at your data and calculate a correlation.

The "correlation coefficient" (also called "r") is a number between -1 and +1. If r equals 0, there is no relationship between the two variables. If r equals +1, then there is a perfect positive relationship between the two variables. When we say a positive relationship, we're not characterizing the relationship as a good thing. In correlation, positive simply means that as one variable increases, the other also increases. For example, you'd probably see a positive correlation (an r that is greater than zero) between the number of hours students in your class spend on homework and their quiz scores. If r equals -1, then there is a perfect negative relationship between the two variables. Again, negative is not a judgment. It simply means that as one variable increases, the other decreases. For instance, you'd probably find a negative correlation (an r that is less than zero) between student absences and quiz scores.

If you haven't already figured it out by now, r falls between -1 and +1, but almost never actually reaches 1 (+ or -). But the closer r is to -1, the stronger the negative correlation between the two variables. Conversely, the closer r is to +1, the stronger the positive correlation.

Two important notes: First, always remember that correlation is not necessarily causation. In other words, just because two variables show a correlation, it does not necessarily mean that one causes the other. We can theorize that an increase in student absences will cause a decrease in quiz scores; but the correlation coefficient alone is not enough to confirm this. Why not? Let me demonstrate with an example: there is a strong, positive correlation (an r greater than zero and close to +1) between the number of storks living in an area in the U.S. and the number of babies born in that area. Do you want to conclude that more storks are bringing more babies? Probably not. In this case, there is actually a *third variable* influencing both of the first two variables—the degree of land development in an area. Storks are more likely to live in rural areas where they can nest close to nature. In the U.S., rural families typically have more children than families living in suburban or urban areas.

The moral is: just because two variables appear to be related, just because they have a correlation coefficient close to +1 or -1, don't assume you've found a cause-and-effect relationship.

The second important note is that correlation should only be performed on data that statisticians call interval/ratio data. These are variables such as quiz scores, age, height, income, temperature, and so on. With this type of data, there is a clear indication of what is higher and what is lower: 70 degrees is warmer than 40 degrees; a 6' tall man is taller than a man who is 5' 6"; a student who scores 90 on a quiz has outperformed a student who scores 60. You cannot use variables such as sex/gender, race/ethnicity, special education classification, or eligibility for free or reduced meals in correlation because no single category for a variable is any better or higher than any other category. These variables do not consist of interval/ratio data.

So that's the theory. Now we can use Excel to calculate it. Using the same spreadsheet of biology quiz scores, absences, and tardies, let's find out if there really is a relationship between absences and quiz scores. First, click on any empty cell on the spreadsheet. (This is where Excel will put r once it's calculated.) Then click the "Paste Function" button just like you did when calculating the average. You'll get the same dialog box as before, and again, click "Statistical" on the left. Then click "CORREL" on the right. Then click "OK." You'll see a dialog box that looks like Figure 5.

Here, Excel is asking you to name the two variables that you want to compare ("Array1" and "Array2"). It really doesn't matter which variable, quiz scores or absences, you put in which array. But for this example, I'll put quiz scores in Array1 and absences in Array2. To do this, click the button for Array1. The dialog box will shrink down and allow you to click anywhere on the spreadsheet. Now, look at Figure 6. Let's say you want to select quiz scores for the lesson "Intro & Scientific Method," you'd click on the "70" (Mark Adams' score) and drag your mouse down to the last "100" (Paula Wyman's score). (Note that you do not want to select the 84.2 average score that you calculated earlier.) Once you've made this selection, click the button to return to the CORREL dialog box. You'll notice that Excel has now filled in the information it needs for Array1. For Array2, repeat the same procedure a to select the students' absences, from Mark Adams' 4 absences to Paula Wyman's 2 absences.

Return to the CORREL dialog box, then click "OK," and in the empty cell you originally selected for Excel to put r, you will see your correlation coefficient. Given the data above, r equals -.60. This tells us that there is a moderately strong negative correlation between quiz scores and absences. In other words, when one goes up, the other goes down.

Try the same procedure, but with different combinations of variables (different quiz scores, or with absences, or number of tardies). You can also combine what we've covered in this article—insert a column between Mitosis and # of Absences or use the blank cells after # of Tardies, then use the function button to calculate each student's average quiz score over the course of the six lessons. Then correlate this result with absences or tardies to see if there is a relationship between overall quiz performance and absences or tardies.

This is just an introduction to the kinds of statistics that you can calculate based on data as readily available as your grade book. By simply exploring your student data, you can glean a wealth of information about your students, their needs, your priorities, and, just like Ms. Logan, you will find analytical steps to success.

Rob Kadel is the founder and a general partner of Kadel Research Consulting, LLC located in Columbia, MD. His firm focuses on the evaluation of educational programs in technology, school reform, and community involvement. With Rob's graduate degrees in sociology and his focus on educational research, the evaluation of technology-supported education became a natural fit.