DATA DRIVEN ACTIVITIES: TEACHING STATISTICS THROUGH MATHEMATICS

Jeffrey A. Witmer
Oberlin College
USA

Real data and statistical techniques can motivate many traditional mathematical topics at the secondary level. Collectively, the important statistical ideas and ways of reasoning can be developed in the context of studying mathematics. The study of formulas, linearity, centers, inequalities, matrices, and logarithms can be embedded in data and statistics and used to lay the foundation for the mathematics and to demonstrate the relevance of statistics to the world. Data Driven Mathematics provides teachers and students with application based activities that makes this happen and that can be used in conjunction with a standard mathematics course or to design a data based statistics course.

"When are we ever going to use this?" is a common refrain among students in mathematics classrooms. It has been said that in teaching mathematics we answer questions that no one has asked. And yet we in the statistics community know that statisticians use mathematics in answering important questions every day. Data Driven Mathematics (DDM) was developed around the idea that we can use statistics to motivate the learning of mathematics, see http://www.pearsonlearning.com/plearn/Feb01/p3.cfm. In this paper I'll briefly outline the DDM materials and how they can be used.

The DDM materials consist of 11 books that roughly match the typical mathematics curriculum taught in the United States in grades 7 through 12. The first book is *Mathematics in a World of Data*. This book invites students to work in cooperative groups to explore dealing with percents and with measurement, constructing graphs, and using scatterplots. There are activities in the book, such as a reaction time experiment in which a ruler (i.e., a measuring stick) is dropped between the fingers of the student, who is to catch it as soon as it starts to fall. A graph is then used to convert distance to time. For example, if the ruler fell 20 cm before being caught, then the reaction time was 0.2 second. Students are then invited to consider the distance that a car travels between the time the driver realizes that the car must stop and the time the driver applies the brakes. As the title of the book suggests, the goal is to get the students to think about mathematics and data in the world around them.

Another book in the DDM series is *Exploring Symbols*. As you might guess, this book is intended to be used as a supplement in an introductory algebra class. The idea of using algebraic symbols to represent real objects, and algebraic formulas to express sensible and relevant relationships, makes a world of sense. Statistics makes this natural: We start with a real-world question and data (that is, numbers in a specific context that gives them meaning). Then we explore and analyze. Later we summarize what we have found, expressing our thoughts in terms that others can understand. In the process of describing what we have learned, we learn it better ourselves - something that teachers experience many times over. Table 1 is an example from *Exploring Symbols*.

Table 1
Motor-Vehicle Registrations (R) Miles Of Travel (M) Motor-Vehicle Deaths (D)

State	R (thousands)	M (billions)	D
Alaska	486	3.8	106
California	22,202	262.5	3,816
Florida	10,232	114.3	2,480
New York	9,780	109.9	1,800
Rhode Island	622	7.7	79

Table 1 shows the number of motor-vehicle registrations (R), the number of miles of travel (M), and the number of motor-vehicle deaths (D) in each of five states for the year 1994.

The question posed to students is how to rank the states in terms of traffic safety. Clearly some kind of rate is needed here. One possibility is to calculate thousands of registered vehicles

per death; another is to calculate billions of miles driven per death (see Table 2). These calculations show Alaska and Florida to have poor records.

Table 2
Thousands Of Registered Vehicles Per Death(R/D) Billions Of Miles Driven Per Death M/D

State	R/D	M/D
Alaska	4,585	35,849,056
California	5,818	68,789,308
Florida	4,126	46,088,709
New York	5,433	61,055,555
Rhode Island	7,873	97,468,354

Students are also asked to consider the typical number of miles driven per vehicle per year for the five states combined. To address this, they need to consider whether it makes sense to find M/R for each state and then average the five answers, or whether they should find the sum of the M's and the sum of the R's and then divide these two totals. Thus, the students need to think about the meanings of the algebraic manipulations that they might want to carry out.

The DDM book *Exploring Linear Relations* helps kids appreciate the use of linear equations in the study of relationships between pairs of variables. I teach at Oberlin College, which enrolls one of the better groups of students in higher education in the U.S. Nonetheless, I sometimes encounter students who have little ability to interpret the slope or intercept of a line. Many of these students have memorized phrases like "the rise over the run" or "the point where the line crosses the axis," but they are weak when it comes to explaining what these things mean. They are also weak in using a linear equation to make a prediction (that is, in thinking about a line as describing a function and in saying what f(x) is for a given value of x). Again, statistics gives motivation and meaning to the study of lines.

Here is an example from the *Exploring Linear Relations* book. Students are given a list of names of a dozen famous persons and are asked to estimate the age of each person. Then they are told the true ages and are asked to make a scatterplot of Y=actual age vs. X=estimated age. The next question is "Which member of the class did the best job of estimating ages?" The line Y=X (*not* the regression line) is the natural line to use to measure how well each student did. Taking the sum of absolute deviations from this line gives an overall measure of accuracy. Later in the book students deal with least squares lines for data such as X=height and Y=weight.

Geometry is a standard subject within high school mathematics and there is a book in the DDM series that fits in with a geometry course: *Exploring Centers*. A major activity in this book involves the study of the mean as center of gravity. Students cut out a triangle or other geometric shape and tape raisins on the vertices. They then balance the shape, for example on the tip of a pencil, and compare the balancing point to the centroid. The idea of a weighted average is studied by putting more than one raisin on some vertices. This eventually leads to the activity of finding the population center of the United States. Along the way, various properties are examined, such as the fact that the three medians of a triangle meet at a single point and that this point is 2/3 of the way from a vertex to the midpoint of the opposite side.

Exploring Logarithms uses sets of data to motivate the use of logarithms. In this book the use of logarithm transformations is shown to be quite helpful in studying relationships. For example, one dataset gives the lengths and weights of 25 alligators. A plot of Y=weight vs. X=length shows marked curvature. However, a plot of Y=log(weight) vs. X=length is linear. Students fit regression lines and examine residual plots to help assess adequacy of the fitted models.

At the upper end of the high school mathematics curriculum one might want to use the DDM book *Advanced Modeling and Matrices*. This book begins with the introduction of matrices as a means of organizing data. The manipulation of matrices through addition and multiplication is presented as a natural way to address sensible questions. For example, the students are asked to consider a rating system to find the best overall city in which to live. They are given a set of ratings for various cities in each of several categories. Then they are asked to consider the effect of using unequal weights when combining the categories to get an overall score for each city. Putting the data into a matrix, with rows corresponding to cities and columns corresponding to

categories, helps one to think about the situation. The weights one wishes to give to the different categories can then be placed into a vector which, when multiplied by the data matrix, gives a set of overall scores. The second half of this book uses matrices in the study of multiple regression modeling. This material can be used in a 12th grade class, for example, after an Advanced Placement Statistics class has taken the AP exam.

Most of the DDM books are intended to be used in connection with a specific course, at a specific grade level. However, the book *Exploring Projects* can be used at almost any level. This book gives a simple introduction to the use of data collection and analysis projects. It discusses how a survey is different from a census or an experiment and goes on to give examples of each of these three ways of obtaining data. Simple experiments are suggested, such as a comparison of two types of popcorn poppers. Mathematics educators are always trying to improve student achievement in mastering skills and understanding how and why mathematics is used. Data Driven Mathematics helps with this effort by motivating important concepts through real-world problems and data.