DATA ANALYSIS: LINKING MATHEMATICS, SCIENCE, AND SOCIAL STUDIES

Jerry Moreno
John Carroll University, United States of America
moreno@jcu.edu

The topic of data analysis is usually found in the mathematics school curriculum and seldom elsewhere. Perhaps this is so because the subject is so often viewed narrowly as a body of techniques and formulas, and not as an investigative problem-solving process that begins with asking a meaningful question for which data will be required to help develop an answer. It is completely natural when the subject of data analysis is viewed in this broader problem-solving process sense to look across the school curriculum, especially in science and social studies, to find contexts in which rich, interesting questions may be formulated. This paper suggests some examples that may be used in the school classroom discussions of the integration of mathematics, science, and social studies through data analysis.

INTRODUCTION

In the workplace and otherwise of everyday life, data analysis, statistics, and probability are pervasive. Making use of them can enrich our lives as they enable us to become better decision makers and become more involved in real world issues from a quantitative point of view. An excellent non-technical publication of essays that illustrates the many venues in which statistics is found is *Statistics: A Guide to the Unknown* (Peck, 2006). In it are found essays on important issues in public policy and social science, science and technology, biology and medicine, business and industry, hobbies and recreation. An excellent online resource on statistical literacy is *StatLit, Statistical Literacy*, www.StatLit.org, which houses an enormous number of papers at the school and undergraduate levels. Much of the material found there requires as prerequisite only an interest in learning what is needed to become a more well-informed citizen. (Milo Schield is the Director of the W. M. Keck Statistical Literacy Project that funded the development of this remarkable website.) The *International Statistical Literacy Project* provides a terrific online repository of international resources and news in statistical literacy, international activities to promote the resources and outreach activities. See www.stat.auckland.ac.nz/~iase/islp/home.

Fortunately there is an increasing awareness of the need for everyone to become statistically literate as witnessed by many sessions in previous ICOTS conferences. Data analysis is found throughout the world's school curricula, typically in mathematics standards as "Data Analysis, Statistics, and Probability," but slowly appearing in science and social studies standards as well under headings such as "Science as Inquiry," "Science as a Way of Knowing," or "Social Studies Skills and Methods." But there needs to be more evidence of the presence of data analysis techniques in the core curriculum standards of social studies and science and the actual implementation of the standards in the classroom.

The tools of data analysis and statistics are typically developed in the mathematics curriculum but often without meaningful contexts, "tools for tools sake" so to speak. Science and social studies contain meaningful real world contexts but lack the quantitative means with which to verify their results. Data analysis links these three core disciplines by providing a problem-solving process that uses the statistical tools developed in mathematics in the rich contexts found in many science laboratory experiments and social studies projects. Data analysis viewed as a problem-solving process is written about very well in the American Statistical Association's *GAISE Report* (Franklin et al, 2007).

The *GAISE Report* gives support and encouragement to state and national Standards writers to include data analysis in various core courses of the school curriculum. The so-called GAISE Model emphasizes that data analysis is a problem-solving process consisting of four steps: formulate a question that can be addressed with data; design and implement a plan to collect the data; analyze the data with appropriate graphical and numerical methods; and, interpret the analysis in the context of the original question. (Franklin et al., 2007, pp. 11-12) It is through the GAISE Model that data analysis links mathematics, science, and social studies and it is through the GAISE

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Model that the scientific method in science and social studies skills for evaluating the credibility and reliability of sources for projects is enhanced and verified.

DATA ANALYSIS ACROSS THE CURRICULUM

One of the best examples that the author has experienced of a school implementing data analysis in all of its subjects was created by St. Vincent-St. Mary High School in Akron, Ohio, USA (http://www.stvm.com/). In an effort to have their students improve their mathematical skills across the curriculum, David Rathz (Headmaster), Pam Godshalk (English/social studies teacher, PGodshalk@stvm.com), and Grace Gorman (foreign language teacher) organized a project prompted by a school self-study and accreditation review in 2004. One part of the project required all school administrators and teachers to attend two workshops on data analysis and probability taught by the author. Once the teachers were reasonably comfortable with appropriate graphical and numerical methods, they coordinated their efforts and introduced quantitative reasoning into their various courses. Students were then required to construct course-contextual projects using data analysis. Pam and Grace offered the following evaluative comments concerning the implementation of data analysis across the curriculum at their school.

> When our school [St. Vincent-St. Mary High School, Akron, Ohio] adopted a school improvement plan that involved instruction in math across the curriculum, many of our staff, especially those in the humanities and the arts, were feeling rather overwhelmed and very much outside their comfort zone of expertise. The strategy to be employed was that all teachers would develop lessons that instructed students in data analysis skills in their particular disciplines. With much initial professional development of the faculty, though, most teachers found that there were numerous ways in which they could engage students in lessons that developed skills in data analysis. Over the four years of implementation of the plan, teachers gained much more confidence in their own abilities to 'teach math across the curriculum,' and student performance on both local and standardized assessments showed significant improvement in math scores. Feel free to contact us regarding our implementation of data analysis across our curriculum. (Pam Godshalk and Grace Gorman, PGodshalk@stvm.com)

The two main important features of the following St. Vincent-St. Mary student examples are 1) that the examples were initiated across and within various departments of the school and not as projects in a mathematics class, and 2) that students were experiencing data analysis as an integral part of every other course that they were taking.

- **Religion course**: A group of students created Operation Christmas Child in which they packaged boxes with gift items for hospitalized children, some terminally ill. For very tight budget purposes that included supply and labor entries, they had to estimate well the number of boxes to prepare as some children were not in need of such gifts for various reasons, some were in need, and some unfortunately would pass away before Christmas.
- **English and Literature courses**: One project investigated the cost of various items such as winter coats, ping pong tables, gas stoves, during the Great Depression of the 1930's as compared to today. Cost data were adjusted for inflation. Questions like "How many hours of work would it take to buy each item in the two periods of time?" were answered. Another project compared the genocide in Darfur Province, Sudan to the Holocaust, while another compared the reasons of suicide from the time of Romeo and Juliet to today.
- **Fine Arts**: Various surveys were taken to compare favorite style of music, favorite instrument, favorite concept studied in Music Appreciation class across grade levels.
- **Foreign Language**: A study was done comparing the price of Big Macs from 15 different countries.
- **Health/Physical Education**: Students investigated the number of football related deaths in high schools across the country for the past 30 years to detect trends. Another project analyzed what discipline/reaction should be used by parents in several simulated parent/child scenarios

such as, “Your 17 year old daughter is pregnant; son is expectant father” or, “Your 5 year old will not stay in his/her bed; wants to sleep with you.”

- **Mathematics**: A group of students developed a model that reflects the height of a plane in flight as a function of its time in the air as it takes off at A and stops at B. They determined an estimate of the average time when in minutes passengers would be permitted to use their computers assuming a condition that the plane has to be above 10,000 feet to do so.
- **Science**: Students compared the rate of gastroschisis births (a type of birth death) versus “normal” births considering the mother’s age, ethnicity, region of country, genetic and environmental causes.
- **Social Studies**: One project conducted a study of voter turnout with respect to age, education, race, employment status, gender, issues. Another investigated immigration and industrial growth in America from the late 19th century to the period immediately preceding WWI.

HISTORICAL CONTEXTS

There are many examples in historical contexts that illustrate how core subjects are linked via data analysis. Some could have dire consequences if the linkage is not understood. Four classical ones come to mind.

- **The Literary Digest and Gallup**: Alf Landon and FDR: Ben Wattenberg is host/essayist in a Public Broadcasting System series entitled The First Measured Century. One session is on George Gallup and the Scientific Opinion Poll with information regarding the demise of the Literary Digest that predicted Alf Landon would defeat Franklin Delano Roosevelt in the 1936 presidential election. Shortly thereafter the Literary Digest went out of business. Gallup predicted the election correctly and began a very lucrative career. The following website describes George Gallup’s beginnings. It contains a teacher’s guide and several lesson plans for the classroom. www.pbs.org/fmc/segments/progseg7.htm
- **The 1970 Draft Lottery**: On December 1, 1969, the U. S. Selective Service performed a lottery to determine an ordering of conscription for military service in 1970. It was based on birthdates. Soon after the ordering was published, questions arose concerning the randomness of the physical selection procedure used. It appeared that the process biased against birthdates occurring late in the year. A different process was used in 1971 that produced an equally likely permutation of birthdates. Using an unbiased process was straightforward and simple to do; demonstrating the bias in the original process was equally straightforward by simple graphing. (Hollander & Proschan, 1984)
- **The Challenger Disaster of January 28, 1986**: A lesson learned from this tragic event was that commercial and governmental pressures coupled with serious lack of communication between departments/organizations, and the lack of understanding that many decision processes can benefit from statistical science are bound to lead to disaster. (Dalal et al, 1989)
- **Authorship of the Federalist Papers**: In 1787-88 essays referred to as *The Federalist* were published anonymously by Alexander Hamilton, John Jay, and James Madison. There were 77 of them trying to persuade New York citizens to ratify the Constitution. The importance of the papers rests in their giving some insight into the thinking of the framers of the Constitution, clearly a topic of importance in political philosophy. For various reasons (one of which was that Madison and Hamilton became political enemies; another is that their styles are very similar), determining who wrote which essay was not easily done. However, there became general agreement on all but twelve of the papers with authorship unclear between Hamilton and Madison. Mosteller and Wallace (1978) applied frequency distribution analysis and probability models to reach the conclusion that it is extremely likely that Madison wrote the disputed essays with one possible exception.

THE DATA LINK IN SCHOOL SCIENCE

In a typical science laboratory experiment, students complete a procedure and determine some sort of answer, for example, the density of a substance. Rarely are class data compared to anything other than an accepted value as found in a reference book. Many science teachers approach laboratory experiments with the attitude that there is a “right” answer and that the

purpose of the lab is to see how close students can come to getting that answer, i.e., the experiment is designed to confirm, not to explore. A program called SEAQL, Science Education And Quantitative Literacy, seeks to foster genuine exploration of data in science laboratory activities that promote a view of science as exploration and modeling, rather than only as confirmation of facts that are already known. For example, how many drops of water can be put on a coin before the water spills over the edge of the coin onto the table? Without any scientific protocol, students' answers will vary considerably. Even with controlling various conditions such as distance the dropper is from the coin, angle the dropper is to the coin, how the dropper is squeezed, which side of the coin is used, variation of results exist. A key goal of SEAQL is to get students to understand the concept of variation, the need to control variables in an experiment, and the need to replicate the experiment. SEAQL frames science experiments according to the GAISE model.

Some basic and enhanced questions for science laboratory experiments in SEAQL include:

- How long does it take an Alka Seltzer tablet to dissolve in water? Does the temperature of the water affect the length of time it takes an Alka Seltzer tablet to dissolve in water?
- How high does soda spurt from a 2 liter bottle of soda when a Mentos candy is dropped in the bottle? What is the effect of dropping k Mentos candies in the bottle? Does the temperature of the soda affect the height of soda spurt? Does the size of the opening of the bottle affect the height of soda spurt?
- How many coins are needed to put in a bottle to sink the bottle? Is the effect of adding coins to a bottle to determine when it sinks a linear process?
- Does compost matter in the growing of dahlias? The author grows dahlias for garden enjoyment and competition. A few years ago, the author helped an fourth/fifth graders in an elementary grade school in Ohio design a randomized experiment involving the growing of their favorite flower, the dahlia. (Typically, students at that school conduct several projects over the summer in which they grow vegetables and flowers, harvest them and sell the "fruits of their labor" to raise money for other science and social studies projects at their school. Often they take flowers to hospitalized friends or to elderly in nursery homes, an activity that informs their social studies curriculum.) More specifically, in one project the author worked with fourth/fifth graders at the school on analyzing "Does compost matter in the growing of dahlias?" See Moreno (2006) for a description of the experiment including the data and their analysis in which the students learned (ex post facto) how important and essential it is to maintain tight control of various variables in scientific protocol in order to obtain valid results and conclusions. The analysis of this experiment can be expanded at the secondary level by using a randomization test. Thanks to Beth Chance and Allan Rossman, an applet exists that tests the hypothesis of no mean (or median) difference between two independent groups. See www.rossmanchance.com/applets, scroll to the bottom left and click on "Randomization Test for quantitative response." The two groups of data may be entered stacked or un-stacked. An entertaining visual of the entered data being randomized can be viewed or turned off. A p-value is computed from which a conclusion is made.

THE DATA LINK IN SCHOOL SOCIAL STUDIES

Generally, it is not as easy to find interesting and meaningful studies in categorical data as in measurement data. The author thanks his friend Michael Kimmel, a retired high school Biology teacher, for the following univariate categorical data activity that cuts across science and social studies and was done by several of his classes when he taught. The observational study involves determining causes of death taken over time in Michael's small hometown on the shores of Lake Erie, Conneaut, Ohio, USA. Details may be obtained from the author.

- Data on a number of variables (e.g., date of death, sex, age at time of death, primary cause of death, secondary cause of death) were available from the town's Bureau of Vital Statistics. They existed back to 1902. One study collected the primary cause of death for random samples taken for four years, arbitrarily taken but spread out over 1902 to the present. The causes of death are very interesting, particularly in the early 1900's for which there were many due to

railroad accidents ("crushed by RR engine," "chest crushed by train," "fell under rr car") and many infant deaths due to stillborn. Of particular interest were some in 1918 such as "cut her own throat" as the primary cause with secondary cause "crazy with flu." This prompts the learning of pandemics, in particular the so-called Spanish Flu, the influenza pandemic of 1918. One way to categorize the many different specific primary death causes is to create categories for them such as: accident/suicide; birthing-related; brain-related; cancer; heart-related; lungs-related; and other. Displaying these categories in a segmented bar graph with the four years on the horizontal axis allows one to see that, for example, accidents/suicide decreased over time whereas heart-related deaths increased dramatically. Recall the reference to railroad accidents in the early 1900's. The next higher category refers to birthing-related that has decreased over time as well. Reasons for the decrease in these two categories prompt studying public policy in, for example, safety regulations in the workplace. Of interest also from a social studies perspective is that Mr. Kimmel's students several years ago interviewed persons who had survived the Spanish Flu, or their relatives. The experience the students had of collecting such real data particularly regarding the impact the flu had on the family's lives was incredibly enriching for them.

- A very interesting study involving census data is the first chapter of *Focus in High School Mathematics - Reasoning and Sense Making: Statistics and Probability* (Shaughnessy, 2009). In it students at Rufus King High School in Milwaukee, Wisconsin compared past and projected census data of the United States, Japan, and Kenya. The data were presented in tables and back-to-back histograms (population pyramids) of five-year age groups, male and female. (The source of the data is www.census.gov/ipc/www/idb/informationGateway.php.) Apart from comparing the three countries statistically, the students' discussed the effect that population shifts in age groups and gender could have on various economic and public policy issues for each country. This is an incredibly rich investigation in statistics and social studies for students to experience. As an example of a population pyramid, figure 1 shows male and female population count in Slovenia for 2010. One conclusion is clear that fewer females are born than males but they outlive males.

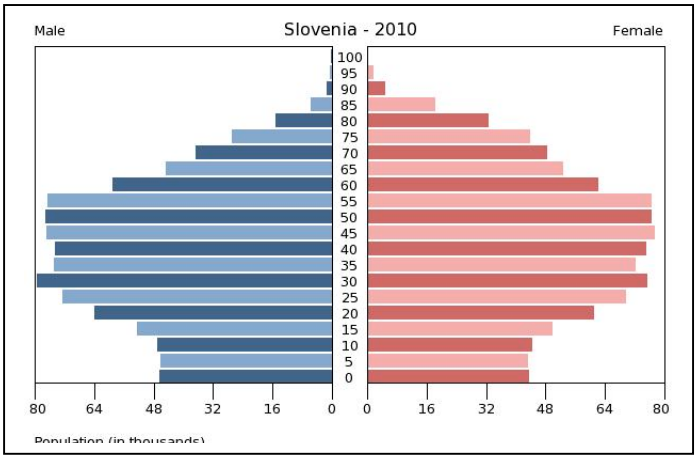

Figure 1. Population Pyramid for Slovenia 2010

- *Making Sense of Statistical Studies* (Peck, 2009) contains fifteen high school classroom-ready investigations covering surveys, observational studies, and experiments. They are excellent activities in real data that show students various contexts in which statistical reasoning is useful. Several cover social study issues. For example, one of the investigations looks at comparing adherence to a curfew time for young people in the United States compared to those in Sweden. Topics for the observational studies include the extent to which people wash their hands after having used a public facility, comparing the contents of different brands of hot dogs, examining if the popularity of first names changes of time, trying to identify a culprit based on shoe size evidence in a crime, exploring seat belt usage over states, and deciding

whether the dimensions of the "golden" rectangle are really most pleasing to the eye. Topics for the surveys cover: characterizing a town by its residents' age, preferences in fast food and clothing and television programs; comparing boys and girls on their interest and participation in sports; conducting a survey on tooth-brushing behavior. The experiments cover: comparing the effectiveness of low-carbohydrate and low-fat diets in reducing weight and cholesterol in obese adults; determining whether listening to Mozart while performing a memorization task helps students remember better than doing a similar task with no music playing; testing whether people have a preference for blue-colored soda. With social networking of various types being very popular, one of the culminating studies involves investigating, primarily for young people, the extent to which the Internet represents a social highway as well as an information one. *Making Sense of Statistical Studies* clearly demonstrates how data analysis ties together topics in mathematics, science, and social studies by providing fifteen excellent investigations for students to analyze real data in meaningful contexts. More information may be found at http://www.amstat.org/education/msss/index.cfm.

- Perhaps the premier activity in social studies that involves students in their culture as well as comparing it to others is in census data. For example, in the United States a national headcount has been conducted every ten years for the past two centuries. There are many entities that depend on an accurate counting of a region's population, e.g., funding, political representation, retail locations. So, a concerted effort is made to guard against undercounts. There are many variables in the census database that students can use for comparison purposes.
- A very interesting and exciting international project that involves gathering real data from school students from participating countries is *CensusAtSchool International*. The project began in the United Kingdom in 2000 joined by Australia, Canada, New Zealand, and South Africa. More recently Ireland, Japan, and the United States have become participants. The data base is accessible at www.censusatschool.com. Through data analysis, students learn about the culture of the students their age in other countries.

CONCLUSION

Developing the underpinnings for a quantitatively literate citizenry will not be based solely in the mathematics curriculum, or in the science curriculum, or in the social studies curriculum. But the linking of these core subjects through data analysis will provide the fundamental structure by which our students will become productive citizens in our societies.

REFERENCES

Dalal, S. R., Fowlkes, E. B., & Hoadley, B. (1989). Lesson Learned from Challenger: A Statistical Perspective. *STATS: The Magazine for Students of Statistics, fall 1989, n2*, 3-6, 20-21.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework.* Alexandria: American Statistical Association.

Hollander, M., & Proschan, F. (1984). *The Statistical Exorcist: Dispelling Statistics Anxiety*. New York: Marcel Dekker, Inc., 121-130 (typo page 128, 374 should be 347)

Moreno, J. (2006). http://www.amstat.org/education/stn/pdfs/STN69.pdf

Mosteller, F., & Wallace, D.L. (1978). Deciding Authorship. *STATISTICS: A Guide to the Unknown (second edition)*, Tanur, J.M., Editor. San Francisco: Holden-Day, 207-219.

Peck, R., Casella, G., Cobb, G., Hoerl, R., Nolan, D., Starbuck, R., & Stern, H. (2006). *STATISTICS: A Guide to the Unknown* (fourth edition). Belmont: Thomson Brooks/Cole.

Peck, R., Starnes, D., Kranendonk, H., & Morita, J. (2009). *Making Sense of Statistical Studies.* Alexandria: American Statistical Association.

Shaughnessy, J. M., Chance, B., & Kranendonk, H. (2009). *Focus in High School Mathematics – Reasoning and Sense Making; Statistics and Probability.* Reston: National Council of Teachers of Mathematics, Inc.