The Statistics Teacher Network

www.bio.ri.ccf.org/ASA/stn.html

Number 58 ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability Autumn 2001

Book Review...

Exploring Statistics in the Elementary Grades Books One and Two

Elementary Quantitative Literacy Project

Bereska, Bolster, Bolster, and Scheaffer
Dale Seymour Publications
ISBN1-57232-345-0

The Elementary Quantitative Literacy (EQL) project was sponsored by the American Statistical Association and funded in part by a National Science Foundation Grant. The National Council of Teachers of Mathematics' *Principles and Standards for School Mathematics* states that "middle-grade students should formulate questions and design experiments or surveys to collect relevant data"; as well as "develop and evaluate inferences and predictions that are based on data." This directs classroom teachers to not only enhance their own understanding, but to provide meaningful activities for their students. With greater emphasis on active learning and an increased need to ensure students have the skills to demonstrate their knowledge, more demands are made on teacher preparation time.

Exploring Statistics in the Elementary Grades: Books One and Two are excellent sources for teachers. Book One is designed with students in grades kindergarten through sixth in mind, while Book Two concentrates on grades four through sixth. Both books are a must for classroom teaches independent of the grade level taught. The books provide a foundation in quantitative literacy. Book One addresses Thinking About Data, Line Plots, Pictographs and Glyphs, Bar Graphs, Probability, Stem and Leaf Plots, and Scatter Plots. Book Two continues to build statistical concepts in a developmental sequence. Topics include Tally tables and Frequency Tables, Two-Way Frequency tables, Line plots, Stem and Leaf Plots, Box Plots, Probability, Scatter Plots, and Time Series Plots.

The lesson plans were field–tested over several years with different age groups in different geographical populations. Developed by teachers and statisticians, each teacher-friendly lesson format includes presenting the question, understanding the problem, gathering and organizing data, describing and interpreting data, drawing conclusions, and additional investigations and projects. Within each lesson plan statistical ideas and vocabulary are boldfaced, defined, illustrated and cross-referenced to Ready Reference pages. These pages define a statistical concept, give a sample problem, list steps needed to organize, describe, and interpret the data. In addition, the glossary contains all boldface terms. Book One contains suggestions for primary and elementary level students.

Among my favorite lessons was one that involved the construction of Nuttie Buddies. With a new plant opening in their hometown,

Also In This Issue...

students are asked to determine the number that could be constructed over a given period of time, whether construction should be completed by a single person or an assembly line, and rates per hour, per day, and per year. Students are given assembly instructions and the need for a uniform product is discussed, quality versus quantity. After a time trial simulation of both individual construction and as assembly line construction, data are collected, and displayed in a stem and leaf plot. Measures of center and spread help students view the variation in performance and make predictions for the number produced per hour, per day, and per year. Factors that can cause the predictions to be inaccurate are also discussed.

It is obvious that classroom teachers designed the lesson plans as they are created in a succinct manner, well presented to equip teachers with enough background to give their students a solid foundation. These two books are a must for your *To Do* list.

Reviewed by
Susan J. Bates
Retired Teacher of the Gifted
sbates@rollanet.org

Classroom Lesson...

Learning from Outliers

By Tena I. Katsaounis
pkatsaouni@aol.com

What are Outliers?

Outliers are observations in a data set that are inconsistent with the rest ("unusual" observations). Outliers are not necessarily "bad" or "misleading" observations. In fact, in some cases they could indicate a surprisingly useful case for further study.

Outliers may arise from different sources of variability in the data such as:

- imperfect collection of data
- inadequacies in the measuring instrument (measurement error), and
- inherent variability (for example, measurements of student performance will reflect the amount of variability in the student population).

A distinction should be made among outliers, extreme values, and contaminants in a data set. Consider a sample of size n data from some population, modeled by a statistical model F. If we order the data from smallest to largest, then the smallest and largest values are extreme values, but may or may not be outliers, depending on how they appear in relation to the assumed population model F. On the other hand, outliers are always extreme values in a sample.

It is important to keep in mind that an outlier is considered as being part of the same population (as modeled by F). In some cases, one might encounter data that are contaminants, i.e., observations that come from another population, which can be described using some statistical model G. Outliers may or may not be contaminants, but, from the statistical analysis point of view, we need to consider outliers as possible contaminants.

The idea of extremeness arises from some form of ordering the data, e.g., smallest to largest for numerical data. A difficulty arises when more than one sample is involved in the data set (multivariate case). This is because there is no unique form of "total" ordering for data for multiple samples (multivariate data), but only different types of "partial" ordering. For example, one can replace the observations of a sample by a scalar quantity, e.g., by the sample average, and order the samples in terms of their corresponding scalar quantity. Then observations that yield extreme values of that scalar quantity would be candidates for being outliers.

Difficulties arise also in the case of multiple samples that exhibit some structure such as in regression, design of experiments, time series, data in the form of direction in a plane or in space, and opinion polls. As an example, consider the regression problem $y = a + bx + e$, where observations x from a sample vary with observations y from another sample in a systematic way which can be described using a statistical model $f = a + bx$ and e is a measurement error. In this case, outliers are less intuitively clear, since an extreme value of either x or y alone does not indicate the presence of an outlier for f. In regression, outliers more likely indicate departure from the assumed model f. It is also of interest to assess the influence of observations on the estimation of f, namely by how much f changes if a particular observation is omitted from the data analysis.

In general, the notion of an outlier has to be determined within the context of the problem under consideration. Outliers in a data set present always an interesting statistical issue. The notion of outliers and ways to handle them can be a very useful subject to teach in a classroom setting. Such a subject can initiate discussion about many key ideas involved in a statistical data analysis, such as quality of data, independence of observations, correctness of assumed model, etc.

How can outliers be detected?

Usually outliers can be detected at earlier stages of the overall data analysis, called exploratory data analysis. Exploratory data analysis is a detailed study of a set of data, which provides a way to explore the structure of the data, and informally study important assumptions. Good statistical practice has indicated that one should always look in detail at the data before doing any kind of formal data analysis, such as computing summary statistics, testing hypotheses, etc.

A major part of exploratory data analysis is visual displays of the data such as histograms, stem and leaf plots, boxplots, pie charts, scatterplots, etc. Graphical techniques can reveal any familiar as well as unexpected features of a data set's behavior such as

- the center of the data (mean, median)
- the spread of the data (variance, range)
- any symmetry in the distribution of the data (bell shaped, skewed)
- more than one high concentration area (modes)
- unusually low or high values (outliers)

Exploratory data analysis also emphasizes the benefits of finding the scale of measurement for the data, that results in simplification in terms of some key features such as symmetry, constant variability, straightness of relationships, etc. In some cases, outlying observations in the original data do not appear as such in rescaled (transformed) data.

Practice has also shown that an analysis of data is not complete without a careful examination of the residuals. Residuals are what remain after a fitted model has been subtracted from the data, residual = data - fit. A graphical representation of the residuals can reveal that:

- a curvature exists in a relationship (need for higher order terms in a model)
- variability is not constant (call for rescaling the data)
- effects are not additive (reexamination of the model)
- outliers are present.

Analysis of outliers

There are different reasons for examining outliers. One might attempt to accommodate them in the data analysis using robust methodology. The main idea of robust methodology is that the effect of small changes in the data should be limited. Thus, one should seek methods that pay more attention to the main body of the data and less to unusual ones.

In other cases, the presence of outliers might indicate that the assumed model does not describe the data adequately, therefore one might decide to modify the original model in a way that such observations will not appear as unusual ones.

Sometimes the purpose of a study is to identify and study outliers because of their practical importance for the problem under consideration.

On might also decide to simply reject the outlying observation(s) and use the remaining data if the statistical analysis of the data with and without the outlier(s) is the same for practical purposes.

Historically there are two approaches in analyzing outliers. One approach is to formally test an outlier. Appropriate statistical tests (discordancy tests, for example) can be used to determine whether an outlier should be kept as being a case of special interest or rejected.

A second approach is using statistical procedures mainly for estimating (and possibly testing) values of the parameters (such as a and b in the regression model f, when the errors come from a known model) that are robust against the presence of outliers. A classical example of a robust estimator for the problem of estimating the center of a population model (location parameter) is the sample median. If the observations are known actually to come from a symmetric (bell shaped) population model than the sample mean will be a good estimate of the center. The underlying population model is typically not known. For this reason, an estimator is sought which performs well for several different types of population models even thought it may not be the best available estimator for any particular type population model.

Note: If the sample size n is an odd number then the middle observation of the ordered set of data is called the sample median. If the sample size n is an even number then any value between the two middle observations is a sample median. In this case, half the sum of the two middle observations is usually taken as being the sample median.

References for further reading:

Barnettt, V. and Lewis, T. (1994). "*Outliers in Statistical Data*" John Wiley and Sons.

Weisberg, Sanford, (1985). "*Applied Linear Regression*" Wiley Series in Probability and Mathematical Statistics.

Classroom Activity...

All Kisses are Not the Same

By Joyce Conway
Tallmadge High School
Tallmadge, Ohio
joyce.conway@tallmadge.summit.k12.oh.us

I use this "tasty" activity after my students have made histograms and know how to calculate the standard deviation. Electronic balances are needed.

The question I pose to my students is which type of Hershey Kiss–plain, hugs, or almond–varies the most in individual mass. I have the students think about the question and write their guesses on a sheet of paper. These responses are collected and the results are tallied. It never fails that the students feel the almond kiss mass vary the most. They reason that the almond inside the kiss will cause the variation in mass of the entire kiss to be greatest. Sound reasoning, don't you think?

Our discussion then moves to ways we can measure the variation in mass. Low and behold, after some prompting one of my students suggests the standard deviation. I reply with asking what shape of a graph is ideal to use the standard deviation. Pause, approximately normal someone questions? Yes. What type of graph is useful to determine shape. That's right, a histogram. Now the fun begins as we look at a sample of each type of kiss.

I divide the students into groups of six consisting of three pairs of people. The next day I have each pair in the group bring a 10-16 oz. bag of hugs, plain, or almond so that each group of six has all three types covered. (I always bring a few extra bags, just in case.) Students use electronic balances to measure each kiss in their respective bag and masses are recorded. Each group makes three histograms, one for each kind of kiss, all using the same scale. The shapes are always approximately normal. The standard deviation is calculated for each kind.

Surprisingly, the results are varied. To the students' amazement, it's the plain kiss that sometimes has the highest variation, although their choice of almond is tops at times. I then ask the students to investigate why their very logical guess before the experiment was done is not always right. One possible explanation is that looking closely at almond and hugs kisses, it appears that they are made using a mold. The plain appears to be made by dropping the melted chocolate to form the kiss. This would be the cause of more variation. I have done this activity many times. Just recently the plain kisses look as though they were also made in a mold. In that case, the almond has the highest standard deviation for obvious reasons.

Try this activity with your students and let me know your results. Thanks.

The New Quantitative Literacy...

Mathematics and Democracy: The Case for Quantitative Literacy

L. Steen, Editor
The Woodrow Wilson Foundation
www.woodrow.org/nced/ quantitative_literacy.html

At the American Statistical Association annual meeting last August, I was given a book entitled *Mathematics and Democracy: The Case for Quantitative Literacy.* I took an immediate interest in it because I had heard that there was a movement afoot that was using the term Quantitative Literacy (QL) in a way somewhat different than the definition of QL that I have been familiar with. To me, QL is a term originally coined in the 1980's to define five volumes of hands-on probability and statistics materials developed by the ASA/NCTM Joint Committee in Probability and Statistics Curriculum for Grades K-12. They were written to help teachers satisfy the NCTM Standards of 1989 regarding statistics and have done a very good job in doing so.

The Case for QL is a product of the National Council of Education and the Disciplines (NCED) that is part of the Woodrow Wilson Foundation. Its goal is to advance a vision that will unify and guide efforts to strengthen k-16 education in the United States. At the heart of its work is a national reexamination of the core literacies – quantitative, scientific, historical, and communicative which are essential to the coherent, forward-looking education all students deserve. The current emphasis is on the **quantitative literacy** segment.

In this article, I will quote liberally without specific reference from *Mathematics and Democracy: The Case for Quantitative Literacy*, edited by Lynn Steen that begins with an article describing the case accompanied thereafter by twelve supporting papers. It is stated that to produce a one-line definition of QL or *numeracy* does not appear to be possible. In the 1930's, John Dewey spoke of literacy as *popular enlightenment* meaning "that which enables people to think for themselves, judge independently, and discriminate between good and bad information, as opposed to the passive literacy of being able to understand instructions and carry out procedures routinely." John Dewey's view is manifested throughout the succeeding discussions that attempt to distinguish *quantitative literacy* from *mathematical literacy.*

The readers of this STN article are clearly aware that practically everyone in the workplace from farmers to lawyers, jurors to the accused, manufacturers to consumers needs to be able to think quantitatively. Regardless of whether we look at Dewey's citizen, or consider today's citizen who is being drowned in information, both are "innumerate" using John Allen Paulos' terminology, and Lynn Steen warns that "an innumerate citizen today is as vulnerable as the illiterate peasant of Gutenberg's time."

So what's the problem? Many states already require that all students have three years of high school mathematics, so time on task should not be a difficulty. Increasing standards may not be the solution either. Many states have a standardized graduation test based on the NCTM *Standards* of 1989 and the more recent *Principles and Standards for School Mathematics* (PSSM) of 2000. Yet how many of our students need remedial work at the college level or try to avoid any major that requires mathematics? How many of us college faculty shake our heads in dismay, as even many of our otherwise capable pre-service teachers admit openly to their not liking mathematics? How can a kindergarten child enjoy numbers as they do, but the vast majority of them lose interest by the time they are middle school age? *The Case for QL* argues that the difficulty is that **we do not teach students what they need to know to function as productive citizens.**

"Quantitatively literate citizens need to know more than formulas and equations. They need a predisposition to look at the world through mathematical eyes, to see the benefits (and risks) of thinking quantitatively about commonplace issues, and to approach complex problems with confidence in the value of careful reasoning. Quantitative literacy empowers people by giving them tools to think for themselves, to ask intelligent questions of experts, and to confront authority confidently. These are skills required to thrive in the modern world." Can you hear John Dewey in that?

If we are thinking that our mathematics programs already do that, then why do we generally agree that our students are so poor mathematically? Why are there so many remedial courses for incoming university students? Might it be that the mathematics that comes so naturally to us, and makes so much practical sense to us, and is so beautiful to us, is not what is natural, practical, beautiful, and most importantly **important** for most people?

QL is not suggesting that students should not know the essence of algebra and geometry. However, QL is suggesting that we deceive ourselves in assuming that students who are thusly mathematically literate through typical school mathematics learning are thereby quantitatively literate. We deceive ourselves in assuming that students can transfer the logical thinking of mathematics, the algorithmic skills of mathematics, the abstractness of mathematics to everyday commonplace applications. They do not, and they cannot, but they must. Our mathematics curriculum is designed to produce *mathematically* literate students, which it is slowly improving upon doing (according to TIMSS and other studies), but it desperately needs to produce *quantitatively* literate citizens. "We need to produce a society for whom quantitative reasoning is as commonplace and demands the same respect as do reading and writing."

So, what is quantitative literacy? "It's not the same as mathematics, nor is it watered-down mathematics. It's not statistics either, although it certainly is driven by data-based statistical reasoning. It is an approach to problems that employs and enhances both statistics and mathematics. Statistics is primarily about **uncertainty** while QL is often about the logic of **certainty**. Mathematics is primarily about a Platonic realm of abstract structures, while QL is often anchored in data derived from and attached to the empirical world. QL is inextricably linked to reality."

And what might QL do that mathematics has not done? From one point of view, mathematics has failed many students, who leave high school with neither the numeracy skills nor the quantitative confidence required in contemporary society. There is no justification to maintain a mathematics curriculum for all students that leaves so many of them functionally innumerate. The mathematics curriculum has historically focused on school-based knowledge which most students do not see the reason to master; QL involves mathematics acting in the real world.

"QL numeracy needs to be learned and used in multiple contexts - history, geography, economics, biology, agriculture, culinary arts - an integral part of all subjects. Mathematics has not the ability to do that. What is needed for life or general school subjects is quantitative literacy; what is needed for education or certain career paths such as engineering and physical science is mathematical literacy."

So, what are the elements of QL? The proposers of QL suggest the following:

Confidence with mathematics (vs. math anxiety)—Being comfortable with quantitative ideas and at ease in applying quantitative methods. For example, naturally comfortable in using mental estimates to quantify, interpret, and check other information. QL numeracy is as natural as ordinary language.

Cultural Appreciation—Understanding the nature and history of mathematics, its role in scientific inquiry and technological progress, and its importance for comprehending issues in the public realm.

Interpreting Data—Reasoning with data, reading graphs, drawing inferences, and recognizing sources of error. Mathematics puts formulas or relationships at the center. QL puts data at the center.

Logical Thinking—Analyzing evidence, reasoning carefully, understanding arguments, questioning assumptions, detecting fallacies, and evaluating risks. Individuals with such habits of inquiry accept little at face value; they constantly look beneath the surface, demanding appropriate information to get at the essence of issues.

Making Decisions—Using mathematics to make decisions and solve problems in everyday life. Mathematical thinking must be as useful and ingrained as reading and speaking.

Mathematics in Context—Notation, problem-solving strategies, and performance standards all depend on a specific context.

Number Sense—Having accurate intuition about the meaning of numbers, confidence in estimation, and common sense in employing numbers as a measure of things.

Practical Skills—Being adept at using elementary mathematics to solve quantitative problems in a wide variety of common situations.

Prerequisite Knowledge—Having the ability to use a wide range of algebraic, geometric, and statistical tools that are required in many fields of postsecondary education.

Symbol Sense—Being comfortable using algebraic symbols and at ease in reading and interpreting them, and exhibiting good sense about the syntax and grammar of mathematical symbols.

What areas should be a part of the curriculum that produces quantitatively literate citizens? As a suggestion, consider the following eight:

Citizenship—Virtually every major public issue – from health care to social security, from international economics to welfare reform – depends on data, projections, inferences, and the kind of systematic thinking that is at the heart of quantitative literacy.

Culture—As educated men and women are expected to know something of history, literature, and art, so should they know at least in general terms something of the history, nature, and role of mathematics in human culture. The aspect of quantitative literacy is most commonly articulated in goals colleges set forth for liberal education.

Education—Fields such as physics, economics, and engineering have always required a strong preparation in calculus. Today, other aspects of quantitative literacy (e.g., statistics and discrete mathematics) are also important in these fields. Increasingly, however, other academic disciplines are requiring that students have significant quantitative preparation.

Professions—As interpretation of evidence has become increasingly important in decisions that affect people's lives, professionals in virtually every field (lawyers, doctors, social workers, school administrators, journalists, chefs, architects) should be well versed in quantitative tools.

Personal Finance—Managing money well is probably the most common context in which ordinary people are faced with sophisticated quantitative issues. It is also an area greatly neglected in the traditional academic track of the mathematics curriculum.

Personal health—As patients have become partners with doctors in making decisions about health care and as medical services have become more expensive, quantitative skills have become increasingly necessary in this important aspect of people's lives.

Management—Many people need quantitative skills to manage small businesses or non-profit organizations as well as to fulfill their responsibilities when they serve on boards or committees that are engaged in running any kind of enterprise.

Work—Virtually everyone uses quantitative tools in some way in relation to their work, if only to calculate their wages and benefits.

Finally, what are the skills of quantitative literacy?

Arithmetic—Simple mental arithmetic, estimating arithmetic calculations, reasoning with proportions, counting by indirection (combinatorics).

Data—Using information conveyed as data, graphs, and charts; drawing inferences from data.

Computers—Using spreadsheets, recording data, performing calculations, creating graphic displays, extrapolating, fitting lines or curves to data.

Modeling—Formulating problems, seeking patterns, and drawing conclusions; recognizing interactions in complex systems; understanding linear, exponential, multivariate, and simulation models, understanding the impact of different rates of growth.

Statistics—Understanding the importance of variability; recognizing the differences between correlation and causation, between randomized experiments and observational studies, between finding no effect and finding no statistically significant effect, and between statistical significance and practical importance.

Chance—Recognizing that seemingly improbable coincidences are not uncommon; evaluating risks from available evidence; understanding the value of random samples.

Reasoning—Using logical thinking; recognizing levels of rigor in methods of inference; checking hypotheses; exercising caution in making generalizations.

So, that's a summary of the *Case for Quantitative Literacy*. I encourage you to read *Mathematics and Democracy* ISBN 0-9709547-0-0. You will be hearing much more about efforts to move the school curriculum away from focusing on that which makes calculus the capstone course, to something else that makes mathematics truly meaningful and accessible to **all** students. I believe that they are on the right track. In particular the pre-university mathematics curriculum must reflect both mathematics and numeracy for all students. Both are essential for life and work, and each strengthens the other. The first invited forum to brainstorm this topic was held December 1-2 at the National Academy of Sciences in Washington DC.

I was honored to have been invited. It was very interesting to hear the mathematicians discuss the book. It certainly seemed to me that a lot of what they were saying in trying to describe quantitative literacy was very much statistical in essence and/or an approach to topics as given in ASA's Data Driven Mathematics modules. I have no hard evidence to offer but I sensed that in my conversations with my math colleagues, several of them did indeed recognize that statistics topics and statistical thinking will be playing a very large role in whatever is done to make mathematics more meaningful to students. Keep in touch with the web site cited in the title of the article to track the development of this quantitative literacy movement.

As a bit of an afterthought but it pertains to the topic being addressed, at a NCTM regional meeting not long ago, I attended a session on the **Core-Plus Mathematics Project** (*http://www.wmich.edu/cpmp*), developed by Christian Hirsch of the University of Western Michigan. I was very impressed with the concept and feel that it might be the type of curriculum that the Case would approve. It is to my mind a very exciting four-course high school curriculum that replaces the traditional sequential Algebra - Geometry - Advanced Algebra and Trigonometry, and Precalculus. It is for **all** students.

Each course features:

- interwoven strands of algebra and functions,
- statistics and probability,
- geometry and trigonometry, and
- discrete mathematics.

The first three courses in the series provide a common core of broadly useful mathematics for all students. They were developed to prepare students for success in college, in careers, and in daily life in contemporary society. Course 4 continues the preparation of students for college mathematics. It formalizes and extends important mathematical ideas drawn from all four strands, with a focus on the mathematics needed to be successful in college mathematics and statistics courses.

Conclusion

As students advance through the grades, they should be linking their knowledge of statistics to other content areas in the social and natural sciences. By doing so, they will develop an understanding of the scientific method and the concepts and processes used in analyzing and interpreting data, a necessary knowledge based in order for them to become able to make critical and informed decisions throughout their lives - as consumers, workers, jurors, citizens.

Currently, as the *Case* indicates, citizens have at most a "supermarket tabloid" understanding of statistics. Instead, they should be equipped with sufficient statistical tools, even if just conceptual and intuitive, to be able to have a general understanding of decisions and conclusions that are based on statistical evidence. If ever there is to be a chance of having a quantitatively literate citizenry, then it is hopeful that future generations will be thus equipped through school statistics or through quantitative literacy vis-à-vis Steen's *Mathematics and Democracy* or through innovative mathematics curricula such as Kirsch's **Core-Plus** program. For the college student, I urge us university faculty to review what we are doing and how we are doing it in our college math courses so that all of our students will truly use the math they learn throughout their lives, and will even carry on quantitative conversations at cocktail parties!

Submitted by Jerry Moreno
John Carroll University
STN Editor

Note from the Editor

Cyrilla Bolster, STN Associate Editor for the elementary level, has resigned her position. I thank her for the work that she has done to bring statistics articles to grades K-5. I wish her well.

If any readers are interested in filling this Associate Editor position, please contact me. Thank You

Keep Us Informed ...

The Statistics Teacher network is a newsletter published three times a year by the American Statistical Associaiton - National Council of Teachers of Mathematics Joint Committee on Curriculum in Statistics and Probability Grades K-12.

Send articles by email to the Editor:

Middle School - Susan Bates
sbates@rollanet.org

High School - Joyce Conway
joyce.conway@tallmadge.summit.k12.oh.us

Editor Jerry L. Moreno
Dept. of Mathematics and
Computer Science
John Carroll University
University Hts. OH 44118
moreno@jcu.edu
fax: 216-397-3033

Statistics Teacher Network
c/o American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415
USA

Let us know your Zip +4!

Non-Profit
Organization
U.S. Postage
PAID
Permit No. 361
Alexandria, VA