The Statistics Teacher Network

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

Number 53 ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability Winter 2000

Book Review...

Interactive Statistics

Martha Aliaga and Brenda Gunderson
Prentice-Hall, Inc., 1999
ISBN 0-13-231036-8

Trends in statistical education over the past several years (see, for example, the articles in Gordon and Gordon (1992), the quotes and discussion in the preface in Rossman (1996), and Scheaffer, et.al. (1996)) include a greater emphasis on active learning, a greater emphasis on the understanding of statistical concepts over implementing statistical recipes (accomplished, in part, with the help of appropriate technology), and a greater emphasis on the use of real data. There are now a handful of good introductory (college algebra prerequisite) statistics texts at the high school AP statistics or introductory college level that incorporate these, and other, reforms. I am pleased to say that the text by Aliaga and Gunderson is an excellent addition to these, and that I adopted it for my introductory statistics class.

The active learning focus championed in Interactive Statistics is clearly seen in the *Let's do it!* activities of the text. These activities, interspersed in the body of the text, "are designed as individual or group projects to be completed in class." The authors view the *Let's do it!* activities "in many ways . . . [to be] the heart of the text." To be fair, these "activities" often involve either just short thought questions or computations. But these may be posed to students as individuals or in groups during class time as understanding checks which actively engage them. Other *Let's do it!* activities involve the students by having them gather and then analyze data. Because of the abundance of *Let's do it!* activities, it is probably not reasonable to use all of these in a semester-long course. As the *Let's do it!* activities are placed at appropriate places within the body of the text, those not covered in class can still be used by students in their reading to reinforce or, in some cases, discover statistical concepts.

Let me give you some examples of *Let's do it!* activities. In a section on random sampling students break into groups of size 8 with at least two men and two women in one *Let's do it!* activity. Treating the group as the population students then estimate the mean number of haircuts per year of the population using a simple random sample and, then, a stratified random sample (random sampling using either a given random number table or a TI-82/83 being carefully discussed in the text prior to this point). Another *Let's do it!* activity has students empirically estimate the chance of winning the prize in the 'Ask Marilyn' problem when using the stick and then the switch strategy. Other *Let's do it!* activities include getting answers to sensitive questions using the law of total probability, simulating the sampling distribution of a sample proportion and a sample mean, and testing whether the proportions of men and women who say pepperoni is their favorite pizza topping differ.

Also interspersed within the text are a number of *Think about it* questions. These questions are apparently intended to be looked at outside

Also In This Issue...

of class as part of the reading assignment of students. The *Think about it* questions were designed to "encourage students to make the leap to the next related statistical concept."

Other unusual features of this introductory statistics text include an opening chapter on statistical inference (population, sample, null and alternative hypotheses are defined on page 4!) and the role of statistics in the scientific method, and early coverage of a number of sampling techniques (e.g., simple random, stratified random, systematic, cluster, and multistage).

For students possessing a TI-82 or TI-83 graphing calculator, the *TI Quick Steps* sections at the end of chapters indicate how the TI-82 and TI-83 graphing calculator may be used to perform various statistical tasks. These *TI Quick Steps* sections comprise perhaps 30 pages of the 676 pages in the text. Occasional, appropriate references are made to the TI calculators in the main body of the text but students need not possess any special calculator to use along with this text. As I did last year, I require my introductory statistics students to buy the TI-83. You may freely download my TI-83 programs - such as marilyn.83p simulating the "Ask Marilyn" game - from the Web site *ftp://ftp.sdsmt.edu/pub/sdsmt/ rwjohnso/*. Further details may be found at *http://silver.sdsmt.edu/~rwjohnso/Ggrant.htm.*

Overall I find the text to be lucid. Concepts are emphasized over formulas and recipes. While students are certainly expected to do some computation, summary statistics are not infrequently provided (sometimes in the form of computer output from Minitab, SPSS, or Excel) to cut down on unnecessary computation. Introductory overviews to chapters or chapter sections have been well thought out and make for nice transitions to more technical material for the student. In Chapter 8, for example, a discussion of formal probability terms, notations, and rules is preceded by a section on simulation. In my own classes I have found it useful to begin any probability discussion with an introduction to probabilities estimated via simulation. Such an approach is also used by Aliaga and Gunderson. I personally prefer using the p-value approach to hypothesis testing rather than the classical approach and, while the classical approach is given some space in the text it is the p-value approach which predominates. I have found students to have a very difficult time with understanding the notion of a sampling distribution of a statistic and applaud the efforts that the authors have made to present this to students through the use of simulation. The exercises at the end of the chapters cover an appropriate range in difficulty. On the low end there is a sufficient number of very straightforward exercises to see if a student understands the big statistical ideas presented and has learned the appropriate statistical language. I find an insufficient number of such exercises in many introductory texts. Many of the exercises, especially those in the inferential portion of the book, appropriately involve data from journal articles, newspapers, or magazines.

I have no serious concerns about the text. On some of the pages where students are to produce graphs or fill in table entries there is insufficient space for the students to do so (e.g., *Let's do it!* 7.16 and Exercise 8.57). In just a few sections of exercises there is a bit too much artificial data. This could be replaced by real data from actual studies (or, if it was real data to start with, a citation indicating the data source). I find my students really enjoy calculations connected with lotteries (e.g., probabilities of winning various prizes and the needed size of a Jackpot to make the game fair in the sense of a zero expected value) and was disappointed not to see any such material along with the discussion of binomial coefficients.

At the time of writing this review I was told that by January 2000 a wealth of supplementary materials will be available at the Prentice-Hall web site *http://www.prenhall.com/aliaga/*. The Comprehensive Instructor's Resource Manual, for instance, indicates how long each *Let's do it!* activity takes and its relative importance. This manual also gives teaching ideas and solutions to all the exercises. Also included among the supplementary materials is a training video. The tape "features interviews with the authors and class testers of the preliminary edition" of Aliaga and Gunderson and shows excerpts from the authors' classes at the University of Michigan. It is a suggested resource for those who are new to interactive teaching methods.

Let me conclude by giving a more detailed accounting of what is contained in the text. Chapter 1 is an introduction to inference and the role of statistics in the scientific method. The scientific method is presented as consisting of the following steps: (1) formulate a theory, (2) collect data to test the theory, (3) analyze the results, and (4) interpret the results. The introduction to each subsequent chapter indicates which of these steps is the focus of the material in that chapter. Classical testing with error calculation and p-values are discussed in Chapter 1. Chapters 2 and 3 concern data collection with a discussion of a variety of sampling methods and experimental versus observational studies (both retrospective and prospective). How to perform a critical analysis of experimental results reported in the press and ethical guidelines for statistical practice are also presented. Chapters 4 and 5 concern graphical and numerical summaries of discrete and continuous data including some discussion on faulty or misleading graphical displays and Simpson's paradox. Chapter 6 looks at modeling continuous variables - with discussion focused on the normal, but including uniform and triangular densities, and briefly at modeling discrete variables - but none of the common "named" mass functions are discussed here. Chapter 7 looks at relationships between pairs of variables. In the continuous case the least squares line is presented along with residuals and the utility of residual plots in identifying nonlinear (e.g., quadratic) patterns. Other topics include outliers, influential points, extrapolation, correlation, correlation and causation, and the effect of linear transformations of data on the correlation coefficient. The chapter concludes with a look at associations between pairs of discrete variables including a look at two-way tables, marginal and conditional distributions, and the interpretation of computer output performing a chi-square test of independence. Chapter 8 begins, as mentioned, with an introduction to probability via simulation using a relative-frequency approach. Enough probability is then presented to develop the addition and multiplication rules as well as the law of total probability and Bayes' rule. (Of late I have skipped any formal statement of Bayes' Theorem in my own classes but have my students compute such conditional probabilities by running a large population through a two-way table and taking appropriate ratios. This informal procedure is briefly mentioned in Example 8.7 of the text (c.f. Truxal (1989)) prior to a formal statement of Bayes' Theorem for a two-set partition.) Chapter 8 includes a brief discussion of random variables and the calculation of the mean and variance in the discrete case. The mean of a continuous random variable is stated to be the place at which the pdf would balance. Discussion of the binomial distribution follows in an appendix. Chapter 9 very nicely presents the sampling distributions of sample proportions and sample means through the use of simulation. Hypothesis tests and confidence intervals (and, briefly, their connection) for a single proportion and a single mean are discussed in Chapter 10. Chapter 11 examines hypothesis tests and confidence intervals when dealing with two-sample problems involving proportions or means. In the two-independent sample case for comparing means the discussion focuses on the equal variance case with only a very brief mention of the unequal variance case. One-way ANOVA is discussed in Chapter 12. The intuition behind the F test as well as its computation is presented. I was pleased to see the assumptions behind the test stressed. I was also pleased to see a brief discussion of the problem of multiple comparisons and a presentation of Bonferroni's method here. The text concludes with chi-square goodness of fit, independence, and homogeneity tests in Chapter 13.

References

Gordon, Florence and Gordon, Sheldon, editors (1992), Statistics for the Twenty-First Century, MAA Notes, no. 26, Mathematical Association of America, Washington, D.C.

Rossman, Allan (1996), Workshop Statistics: Discovery with Data, Springer-Verlag, New York, NY.

Scheaffer, Richard; Gnanadesikan, Mrudulla; Watkins, Ann; and Witmer, Jeffrey (1996), Activity-Based Statistics: Instructor Resources, Springer, New York, NY.

Truxal, John (1989), Probability Examples, Monograph Series of the New Liberal Arts Program, Research Foundation of State University of New York, Stony Brook, NY, pp. 37-43.

Reviewed by Roger Johnson
South Dakota School of Mines and Technology
Rapid City South Dakota
rwjohnso@silver.sdsmt.edu

Statistics for the Elementary Classroom: Sorting and Categorizing

Cyrilla Bolster
bolstereducation@hotmail.com

Ours is a statistical society, and to understand the data explosion which impacts on all aspects of our lives, knowledge of statistics is essential to the experience of the younger population of students.

The National Council of Teachers of Mathematics has outlined the scope and content for the Probability and Statistics strand in the Standards. It is recommended that the teaching of Statistics begin with students as early as the first grade. How can this be accomplished - by making the presentation of material as appealing, motivating and grade-appropriate as possible. Many statistical notions are intuitive and can be introduced at a very early grade level even though the formal terminology might not be introduced until later grades. One of them is the concept of sorting and categorizing.

Learning the Process of Identifying an Attribute:

Before students can sort and categorize items, they must have an understanding of what it means to group items according to one attribute. The game called "Gate Keeper" can be used for this purpose. The game is played with a set of attribute blocks consisting of shapes (triangles, circles, squares) that are of different colors (red, yellow, blue, green), sizes (large, small) and thicknesses (thick, thin).

The game begins by giving each student an attribute block. All participants stand in a line, holding their attribute block and walking up to the "gatekeeper." The adult, who is the gatekeeper, has one attribute in mind (perhaps, "Yellow"). Participants file past the Gatekeeper who says "Yes" to some (those holding yellow) and "No" to others (those holding any other color than yellow). The Gatekeeper may want to make an exaggerated movement with the arm to simulate the bar used at a toll booth. The "Yes" participants are admitted through the gate, and form a line, holding their attribute piece for others to see. The "No" participants turn away from the gate, and form a line, holding out their attribute pieces for the others to see. The Gatekeeper is the only one to speak; the participants play in silence. At the end the adult asks: "Can you see what was needed in order to pass through the gate?" Students identify that the piece had to be yellow. The adult asks: "What did we use for sorting: color, shape, size or some other quality?" Students should recognize that the pieces were sorted by color. Then, indicate the "No" group and ask: "What did this whole group have in common?" Students need to understand that the whole group was "Not Yellow."

Play again. Students keep the same attribute block, and reform into a single line. They come to the gatekeeper who has an attribute in mind, (possibly "Square"). Once again, students file past and are admitted only if they are square. The game is silent, except for the gatekeeper's "Yes" or "No." This allows the participants to use their senses to observe to arrive mentally at a conclusion. Students should identify that the sorting was by square shape. The other group not accepted were "Not square." Continue to play until all students grasp the concept of sorting by one attribute.

Testing Understanding:

Admit several students (perhaps those holding large shapes). When the next student comes to the gate, ask: "Am I going to let you pass?" By observation, each student should be able to correctly identify whether he/she is holding the piece containing that same attribute. If individuals are unable to say, then ask the group of students: "Am I going to let this person pass?" Each time, at the end of the game, ask: "What did we use for sorting: color, shape, size, or some other quality?" (Note: Avoid using thickness as a sorting factor in this game, as it cannot easily be identified from across the room.)

Lesson Title: Vary, Vary Interesting Things

Students of all ages in primary and early intermediate grades love this lesson! It involves handling "things," and the items can vary to suit the interest of the groups.

The objective of this lesson is to have students analyze the items contained in each bundle, and sort them into sets using attributes of color, shape, size, texture, use, etc. This level of thinking involves classification, a notion used in constructing bar graphs, histograms, Venn diagrams, etc. Students will represent their groupings by representing them on paper. Younger students may draw pictures. Older students may choose a Venn Diagram, pictograph, or bar graph.

Some preparation for this lesson is needed ahead of time, and that involves gathering the "things" to be classified. The suggestions are endless: miscut keys from the hardware store or old keys from around the house; button collections; Matchbox-size toys; figurines; writing implements (pens, pencils, mechanical pencils, colored pencils, thin magic markers, etc.); baseball caps; seashells of all types (small ones varying in shape and color are best); wallpaper and fabric swatches; cut-out catalogue pictures; pebbles that represent a variety of rock types.

Once the materials are collected, select a "bundle" of 20-40 items from within a group and place them in a Ziploc bag. The items should be similar, but have enough variety of color, shape, size, texture or material that students can easily identify how to organize them into subsets. There should be enough bags of items that groups of 3 to 4 students will have a collection from which to choose. For typical class size, there should be 8 to 10 bundles of items. There can be bags of "repeat" items, such as two bags of keys.

At the beginning of the lesson, tell students that it is interesting to see how people organize things. You are interested in having them sort a collection of objects into 2, 3, or 4 piles. You want them to describe the quality that each item has in order to be grouped together, and why those in the other piles can't be in that group. Ask them to represent their groupings on paper. Remind them to add labels and to title their work. Students in younger grades may arrange the actual items on different colored sheets of paper. An adult can write the label/ description used, and print out the title for their "graphs." Older students may choose to make a pictograph, bar graph or Venn diagram.

Give the students a time limit to complete the activity. At the end of that time, have them show their groupings to the rest of the class and describe how they categorized their items. Have students challenge each other, question each other, etc. Once all presentations are finished, ask the students to go back to their materials, and regroup them according to completely different criteria. You will find that on the first attempt, they go for the most obvious attribute - most often, color. On the second attempt, they often choose a more subtle attribute, like "texture" or "use."

This activity is valuable for dialoguing and participating in group work. The teacher can use the data displays as a diagnostic tool: Did students choose the Venn Diagram? If so, did they correctly represent the intersection as items that contain two distinct attributes? Did they correctly use bar graphs? Did they label each category? Provide a title? Add a legend? Correctly scale the bar graph? Include the zero (origin) at the baseline of the scale? Did they mistakenly choose a line graph representation? Did some attempt a pie chart? Did students correctly partition the pie chart? The teacher can assess the graphic displays for accuracy or error, and determine which future lessons need to be presented on graphing skills.

Once the construction of the displays is completed, the most important part of the lesson takes place: describing and interpreting the information. Questions should be directed to the whole class, since the graphic representation should be clear enough for all to read and interpret easily. Ask questions such as: What was the sorting factor (attribute) used? Why did you pick this way of sorting? How many categories are there using this variable? What objects were the most? The least? What other ways could the objects have been sorted? What made categorizing materials in this way interesting? Did any other group present a display that used the same attribute for sorting? Are there any displays that look similar?

A follow-up assignment can be given in which the students write a note to one of the groups in which they give feedback on their display. They can congratulate the group for a clever, descriptive or colorful display and they can go on to make observations or suggestions about the graphic display, or pose questions to the group.

The graphs can be posted on the bulletin board, or set up learning stations with different materials, inviting individual students to sort and represent their materials using a visual display.

How does this lesson fit within our study of statistics? When compiling data, very different pictures can be obtained, depending on the variable selected. Facets of the data are revealed with each changing picture. A wide variety of "pictures", or graphs can be used to display data from different variables. This is a mature notion about manipulating statistics, but even young students are capable of insight into the way data can be displayed simply, and then morphed into a more complex picture.

The study of statistics in the elementary classroom is exciting. If you are a teacher, share lessons with a colleague. If you are not a teacher, this may be an area of expertise you can offer to your neighborhood school. Consider it.

Editor's Note: There are two books that accompany the NCTM/ASA statistics program for elementary grades called Elementary Quantitative Literacy (EQL). Cyrilla is one of the co-authors of the books. Book One is primarily for grades K-6, and Book Two for grades 4-8. They are published by Dale Seymour. "Vary, Vary Interesting Things" is in Book One, complete with lesson plan and extension ideas. For information on how to organize a workshop for elementary teachers, contact Judy Dill, ASA Center for Statistics Education, *judy@amstat.org*, or the Editor.

Statistics in the Classroom...

About Factorials

Donald Dessart
The University of Tennessee
ddessart@utk.edu

Students and teachers of probability and statistics frequently calculate permutations or combinations. Years ago one used formulas involving factorials to determine these numbers. Today, there are keys on most calculators that calculate permutations and combinations at the press of a button. Nevertheless, many teachers still have their students use the formulas involving factorials to provide a deeper understanding of the mathematical process. For example, using the TI-83, the approximation of 50! (50 factorial) is shown to be the number 3.04140932 E64. An interesting question arises. How many of the 64 digits to the right of the digit 3 are terminal zeroes? This article will discuss two different ways to calculate the number of terminal zeroes in m!.

A straightforward approach:

Consider a smaller factorial, 10!, and determine the number of terminal zeroes there are. Write "10x9x8x7x6x5x4x3x2x1" and use a calculator to multiply the numbers, one by one, checking each product to be certain that it is not an approximation. The result is 3 628 800 and thus has two terminal zeroes. Thinking about this, one can conclude that a terminal zero was obtained when multiplying by 10 and another terminal zero when multiplying by 5.

Similarly, we reason that zeroes for 20! will be produced when multiplying by 5, 10, 15,and 20 to yield four terminal zeroes. Returning to the original problem, 50!, one would reason that there are factors 5, 10, 15, 20, 25, 30, ..., 50 and conclude that multiples of 5 (actually the product of 5 and an even number) each produces a terminal zero, and moreover the multiples of 5^2 (i.e., 25, 50, 75, 100) each produces an additional terminal zero. So, 50 has 10 multiples of 5 ($50 \div 5$) and two of these are multiples of 25 ($50 \div 25$) and therefore 50! has 10 + 2 or 12 terminal zeroes.

In general, then, the number of zeroes in m! is $[m/5] + [m/5^2] + [m/5^3] + \ldots$ where [ ] indicates the greatest integer function.

An approach of finesse:

A similar problem can be found in the Hungarian Problem Book II, (p. 100). This book provides a very nice theorem concerning terminal zeroes attributed to the mathematician, Adrien Marie Legendre (1752-1833). An adaptation of this theorem states: *"If m is a positive integer, the number of terminal zeroes in m! is given by $\frac{m-s}{4}$ where s is the base ten sum of the digits of the representation of m in base five."* For example, 50 can be written as $2\,(5^2) + 0\,(5^1) + 0\,(5^0)$, so $50_{ten} = 200_{five}$, and therefore $s = 2+0+0 = 2$.

Finally, $(m - 2)/4 = (50 - 2)/4 = 12$; so, there are 12 terminal zeroes in 50!. The proof of this theorem is also found in the Hungarian Problem Book II (p. 101).

Martin Gardiner in the Scientific American (1967) describes how computers determine factorials. For example, 100! has 158 digits in its product of which the last 24 digits are zeroes. Some computer experts have even had factorials printed in the form of Christmas trees for their Christmas cards.

Mathematicians developed formulas to give approximations to factorials, long before computers were invented. James Stirling (1692-1770) developed such an approximation formula for m!. He showed that the ratio of m! and $(2\pi)^{1/2} m^{(m+1/2)} e^{-m}$ approaches one as m approaches infinity (Feller). The percentage error in using this formula as an approximation for m! decreases steadily as m gets larger, and it is remarkable that the formula involves two well-known transcendental numbers, π and e.

References:

Feller, William. An Introduction to Probability Theory and Its Applications. New York: John Wiley and Sons, 1968, 3rd ed.

Gardner, Martin. "Mathematical Games." Scientific American (August, 1967): 104-8.

Rapaport, Elvira. Hungarian Problem Book II. Washington, D.C.: Mathematical Association of America, 1963.

Statistics in the Classroom...

Bringing "Life" to Piecewise Defined Functions

Murray H. Siegel
Sam Houston State University
mth_mhs@shsu.edu

A standard part of the curriculum for most algebra and precalculus courses is the topic of piecewise defined functions. Typically, a function such as $f(x) = x^2 + 1$ (for x less than 2) and $f(x) = 2x + 3$ (for x greater than or equal to 2) is introduced to investigate piecewise defined functions. Most students fail to see the importance of this type of function, except that it is on the next test. Using interesting data from the real world, the teacher can make the study of this type of function meaningful and useful. The following are the life expectancy figures for the United States for the census years from 1900 through 1990. The life expectancy represents the median of the age at death of all Americans born in that year.

Year of birth	Life Expectancy
1900	47.3
1910	50.0
1920	54.1
1930	59.7
1940	62.9
1950	68.2
1960	69.7
1970	70.8
1980	73.7
1990	75.4

A scatter plot is shown to the class utilizing the TI-83 view screen. Figure 1 is that scatter plot.

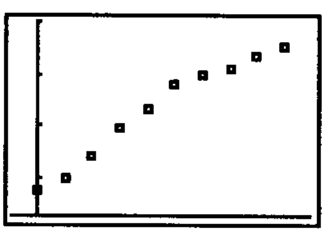

Figure 1

The x-axis shows the number of years since 1900 (from 0 through 90). The y-axis is the life expectancy, from 40 through 80. The class will recognize that life expectancy has been increasing throughout the twentieth century, but the rate of increase slowed in the latter part of the century. In fact, it appears that the change in the slope of the model happened in 1950.

Using the coordinates (0, 47.3) and (50, 68.2), a slope of .42 is computed. The y-intercept is estimated to be 45 from the scatter plot. Using the equation editor of the TI-83, the class establishes that a reasonable fit equation for the years 1900 through 1950 is $y = .42x + 46$. Figure 2 shows the scatter plot and the fit line.

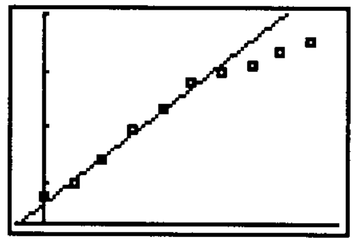

Figure 2

Using the data for 1950 and 1980, a slope of .18 is computed. After a bit of experimentation the class agrees that a reasonable equation for a fitted line for the second half of the century is $y = .18x + 59$. Figure 3 shows the scatter plot and the graph of this model.

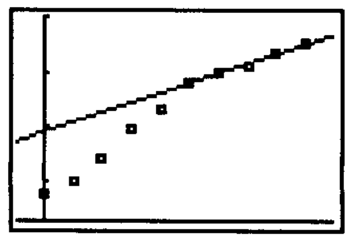

Figure 3

The class can discuss the meaning of the two slopes, in terms of the data. They can also discuss the meaning of the y-intercepts. The more interesting discussion is the reason for the significant change in the slope. If your school or college has a "writing across the curriculum" program, this is a perfect opportunity for students to write in a mathematics class. The author has used these data in numerous high school and university classes and the rationale for the change in the slope has included Viet Nam, World War II, medical technology, drugs and many other conjectures. The correct answer is NOT important. What IS important is that students are talking with passion about a mathematical model. Most recently, the author has used this activity in his Functions and Graphs class. This is an undergraduate precalculus class for elementary education majors who are "specializing" in mathematics. At the conclusion of the activity, not a single student asked why we needed to learn about piecewise defined functions.