The Statistics Teacher Network

Number 45 ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability Spring 1997

Statistics in the Classroom

Teaching Data Analysis to Primary Teachers

Three years ago I was asked to develop and teach a course on data analysis to primary teachers, as part of the NSF-funded Primath Project. The goal of this short, summer course was not to teach the teachers how to teach statistics, nor to teach them by using curriculum materials developed for elementary or secondary students. Instead, I was to teach the basic material on descriptive statistics found in a typical introductory, college-level statistics course. These teachers wanted a solid background in data analysis that would be deeper than the level of material they would need to teach their students.

The course was designed to meet each morning for three and one-half hours, for a two-week period. Afternoons were devoted to other components of the Primath project, but allowed some time for the teachers to make connections between the work we did in the mornings and the types of activities they might use in their own classrooms.

Prior to the first class I asked each teacher to write a short paper for me, consisting of their candid responses to the probe, "Statistics and Me." Reading these papers helped me understand how these "students" would approach my course, and it provided a good introduction to the people that I would be teaching during the summer. I learned that some students were excited about the opportunity to learn statistics, and that some had had negative experiences in a college statistics course. Acknowledging that many of these primary teachers had the same types of apprehensions found in typical college students enrolled in their first statistics course, I was determined to design a course that would provide them with a positive, non-threatening learning experience. I wanted to build the course on active-learning activities and to provide strong connections between statistics and their use in the real world.

I decided that the course should include the following components:

1. Daily discussions of current statistical issues in the news, including the examination of graphical representations of data that appear in the media.
2. The regular use of computer technology to simplify data analysis and to illustrate statistical concepts. All teachers were asked to bring a Macintosh computer from their schools to the class for the entire two-week period.
3. An individual or group project where teachers would integrate and apply their statistical skills and knowledge to solve a problem of interest to them.
4. Daily activities involving real data, either gathered from the teachers, or from already existing (and interesting) data sets.
5. Short segments from the COMAP video series, *Decisions Through Data* (DTD), to illustrate real world applications of statistical ideas.

When I teach a typical introductory statistics course to college or graduate students, I use a

Also In this Issue...

text that has good explanations and examples, and assign sections for students to read before coming to class each day. However, in this summer course for teachers, I was advised that this procedure was not feasible. I decided to adopt a different approach, which was to have students "discover" concepts by working through detailed activities. I was able to obtain a pre-publication version of Allan Rossman's *Workshop Statistics* text which provided this discovery approach and contained many wonderful data sets and activities.

I developed the following schedule of topics for the course:

- Data, Variables, Distributions
- Displaying and Describing Distributions
- Numerical summaries: Measures of Center
- Measures of variability and boxplots
- Comparing distributions
- Looking for relationships involving bivariate data
- Describing bivariate relationships: correlation, regression
- Categorical relationships and Simpson's Paradox

We began our first session with several activities. First, I gathered some data on the teachers' travels (both within the U.S. and abroad), and used these data to illustrate some basic statistical terms and ways of presenting data. Next, teachers were asked to design some questions for a class survey so that different types of interesting data could be gathered and analyzed during our course. We quickly assembled the survey, gave it to the teachers to complete, and entered the data into the Datascope spreadsheet. During this first class we also watched the *DTD* video, *What is Statistics?* to provide an overview of the different types of statistical applications. At the end of class, teachers were assigned to look for examples of statistics in the news: graphs, reports of research, polls, etc., and to bring them to class the next day.

On the second day, we discussed the examples from the news articles brought to class and examined the many different disciplines that depend on statistics (e.g., business, sports, medicine, government, etc.). We began to examine the shapes of distributions resulting from variables in our class survey, with teachers working individually or in pairs on their computers. Later in this class we viewed the *DTD* video on histograms, and then discussed details on how to design and complete the assigned project. Teachers were asked to try to identify a problem of interest to investigate and to decide with whom they would like to work.

Soon the class developed a regular, comfortable routine. We started each day by talking about articles in the news. Then, we looked at the day's new activity, which sometimes began by collecting data from the teachers. After an overview of the day's topic and activity, the teachers would begin to work on the daily activities, often discussing the questions and helping each other. I would circulate among the teachers, asking and answering questions. At different times, the activities instructed them to move to their computers, and would guide them through an examination and analysis of data. About halfway through the morning we would take a break, then watch a video segment, discuss it briefly, and return to the activities. When finished with the activity, we would assemble together as a large group, to summarize some of the key statistical ideas. The teachers would spend the remaining class time working on their projects.

On the last day of the workshop, teachers brought their completed projects to class, displayed on colorful posterboards. We set up the posters around the room and then listened while each teacher or group of teachers took turns presenting the story of their project. These posters were not only beautifully constructed, they provided detailed analyses of statistical data. Some examples of teachers' projects were: a study of the nutritional content of different types of hot-dogs, a comparison of prices at sample of grocery stores, a survey of teachers' driving times and gas expenditures, and an experiment involving a test of visual perception. The teachers were quite proud of their projects and what they represented about the teachers' newly developed proficiency in using and interpreting descriptive statistics.

Since that first workshop I have given additional workshops to elementary teachers and will soon give a workshop to middle school teachers. I've discovered how much I enjoy teaching data analysis to teachers and look forward to continued involvement with precollege teachers.

Joan Garfield
University of Minnesota
Minneapolis, Minnesota
jbg@maroon.tc.umn.edu

Efficient Storing of Statistical Data in the TI-83

Many statistics courses, including advanced placement, are using the TI-83 calculators for analyzing data. While many courses will also use student versions of statistical software residing on computers, graphing calculators provide a consistent tool which students are most likely to have both in class and at home.

Data sets used in the classroom come from a variety of sources. Some are supplied on disk from publishers for their textbooks and activity books; others can be downloaded from the internet or provided directly by students doing surveys or projects. However they are obtained, rarely are they in a form that will load directly into the statistical lists of the TI-83 calculator.

Since there will be many more data sets than can be stored in a calculator at any one time, the data sets must eventually be stored on a computer. Sets can then be downloaded to a calculator as necessary using the TI-Graph Link.

A Better Way to Store Data

Before transforming data into TI-83 formats, an important decision needs to be made about how the data sets will be stored in the calculator. Students will need to store more than one data set in a calculator so that a reasonable number of sets will be available for a few nights homework or to use class time efficiently.

The traditional method of storing data sets in a calculator for statistical analysis is to use lists which are then grouped into one file by the TI-Link software. However, this method has the following disadvantages:

- Data sets which each contain a list with the same name cannot be maintained in the calculator without changing names.
- If the data in the calculator are corrupted, there is no way to reconstruct them with out reloading from either a computer or another calculator.
- The TI-83 software for the TI-Graph Link does not allow viewing of grouped lists before downloading to the computer. (Groups must be ungrouped before viewing.)

However, these disadvantages can be overcome by storing each set of data in a separate program as explained below. An additional advantage of storing data in programs is that it takes less storage.

Storing Data in Programs in a TI-83

Consider the distance and fare data for the airline activity in Rossman's *Workshop Statistics* book. The following program stores these data into two lists of data named DIST and FARE.

```
PROGRAM:AIRFARE
SetUpEditor DIST,FARE
{576,370,612,1216,409,1502,946,
   998,189,787,210,737} STO DIST
{178,138,94,278,158,258,198,188,98,
   179,138,98} STO FARE
```

While programs can be entered directly into the calculator in this way, the process is hard to read on the screen. There is another way to enter the actual data that makes the whole process easier to view.

A. Put data into lists called DIST and FARE.

   Press STAT
   Enter 5:SetUpEditor
   Press 2nd ALPHA and enter DIST,FARE
   Press ENTER
   Press STAT
   Under EDIT Press 1:Edit
   Enter the data in list DIST
   576,370,612,1216,409,1502,936,
   998,189,787,210,737
   Enter the data in list FARE
   178,138,94,278,158,258,198,188,98,
   179,138,98

B. To create a new program called AIRFARE:

   Press PRGM
   Select NEW and press ENTER
   Type AIRFARE and press ENTER

   Press STAT. Select 5:SetUpEditor
   Press ENTER.
   Press LIST. Select DIST from the list of names.
   Press "," (the comma key)
   Press LIST. Select FARE from the list of names and press ENTER.

Note: "SetUpEditor" forces only the lists DIST and FARE to be displayed in STAT EDIT after the program is executed.

C. The following steps transfer the actual data from the lists to the program:

Press RCL.

Press LIST. Select DIST from the list of names.

Rcl LDIST will appear at the bottom of the screen.

```
PROGRAM:AIRFARE
:SetUpEditor DIS
T,FARE
:



Rcl ʟDIST
```

Press ENTER. The actual data will be entered into the program.

```
PROGRAM:AIRFARE
:SetUpEditor DIS
T,FARE
:{576,370,612,12
16,409,1502,946,
998,189,787,210,
737}→DIST
:
```

Press STO. Press LIST. Select DIST from the list of names. Press ENTER.

D. Repeat the process above for the list FARE.

E. Use a TI-83 Graph Link to save the program on a computer.

Transferring the Data to TI-83 Lists

A. Use a TI-83 Graph Link to download programs with the required data to a TI-83.

B. Whenever a data set is needed, execute the program which contains the needed data.

C. If by design or error a data set is altered, it can be reset to its original form by executing the appropriate program again.

Notes

The TI-83 can store a large number of data sets without running out of memory. However executing a program does not clear other lists from memory. It is suggested that un-needed lists be cleared from memory from time to time using the 2:Delete...option found by pressing the MEM key.

A great deal of time can be saved by grouping all the program files for a given assignment using a TI-83 Graph Link. Then students need only to make one selection to download an entire assignment.

My AP Statistics class has been using this process since September for the extensive data sets involved with Allen Rossman's wonderful book *Workshop Statistics*. I highly recommend it.

Albert Coons
Buckingham Browne & Nichols School
Cambridge, Massachusetts
alcoons@aol.com

SEAQL Techniques in a Physics Classroom

A common thread linking all science courses, is the importance of students engaging in hands on activities, in which data are collected and analyzed. What experiments are carried out, of course, depends upon the subject matter and grade levels involved, but how they are analyzed can be similar for all grade levels. This can be done using the statistical techniques advocated by The Quantitative Literacy Project.

Background

During the Summer of '96, I attended an NSF sponsored workshop entitled, "Science Education and Quantitative Literacy (SEAQL)," at San Jose State University. The objective of this workshop was to show science teachers how to make science laboratory work more meaningful to students, through the use of simple statistical techniques.

The SEAQL approach to laboratory experiments focuses on class results, rather than individual results, by having students pool their data. Then by using stem and leaf plots, box plots, and median fit lines, they analyze this information and draw conclusions about the experiment. This approach encourages critical thinking, and makes the laboratory experience more meaningful, because it models more closely how scientists really work.

I now require my students to use these techniques in all of their laboratory experiments, and I am encouraged, both by their positive response to working with class sets of data, as well as by the results obtained. In the past, errors less than 10% in physics experiments were usually considered excellent, however, this year errors were often below 5%. Following is a case study of a laboratory investigation that was carried out to study Newton's 2nd Law.

Case Study

The setup for this investigation consisted of a low friction cart on a horizontal surface, accelerated by a mass hanging over the edge of a table, using a fixed pulley, as shown in Figure 1.

Two physics classes, Period 1 and Period 4, elected to keep the hanging weight constant for all trials, and vary the total mass for each trial. Newton's 2nd Law would then be expressed as

$$a = Wm^{-1}$$

where a is the acceleration, W is the weight of the hanging mass, and m is the total mass of the system. A graph of acceleration versus the reciprocal of the total mass should be a straight line through the origin, with a slope equal to the hanging weight.

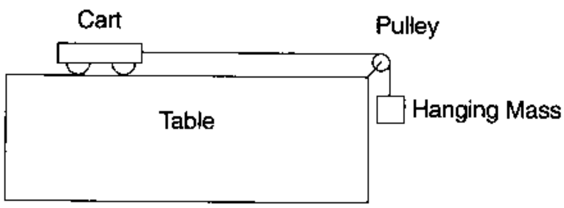

Figure 1—Experimental Setup

My Period 6 class elected to keep the total mass constant for all trials, and vary the hanging weight for each trial. Newton's 2nd Law would then be expressed as

$$a = m^{-1}W$$

A graph of acceleration versus the hanging weight should be a straight line through the origin, with a slope equal to the reciprocal of the total mass.

Each class actually performed this experiment twice. For the first one, a standard spark acceleration timer was used to collect displacement versus time data. In the second experiment, a Texas Instrument CBL unit was used with an acoustic motion detector. For each trial, the acceleration was constant, therefore the displacement function, y(t), could be expressed as

$$y(t) = \frac{at^2}{2} + v_0 t + y_0$$

where v_0 is the original velocity, and y_0 is the original displacement. Since any data point could be chosen as the initial point, y_0 can be set equal to zero, and this equation can then be expressed as

$$\frac{y}{t} = \frac{at}{2} + v_0$$

Each student graphed their data as y/t versus t, and, using their TI-82 calculators, found a median fit line for the scatter graph. The experimental value for acceleration was equal to twice the slope of this line, and individual results were then organized into a set of class data. Again, a median line fit was determined, and the results for the three classes are summarized in the following two tables.

As can be seen from these tables, the errors in four of the six cases are less than 5%. One class actually had errors less than 1%, for both experiments. Compared to them, the errors of 13% and 17% appear excessive, however, the surprising thing is that up until this year,

Table 1—Spark Acceleration Timer

Periods	Equation	Accepted Value	Experimental Value	Error
1	$a = 1.17m^{-1} - 0.160$	F = 1.03 N	F = 1.17 N	13%
4	$a = 1.49m^{-1} - 0.0110$	F = 1.52 N	F = 1.49 N	2%
6	$a = 0.792W - 0.0638$	m = 1.27 kg	m = 1.26 kg	0.8%

Table 2—CBL

Periods	Equation	Accepted Value	Experimental Value	Error
1	$a = 1.00m^{-1} - 0.0283$	F = 1.04 N	F = 1.00 N	4%
4	$a = 0.456m^{-1} - 0.0326$	F = 0.546 N	F = 0.456 N	17%
6	$a = 1.52W - 0.0356$	m = 0.651 kg	m = 0.656 kg	0.8%

these errors would not have been viewed as too disturbing. In addition, the students noted that the median fit line did not pass through the origin as predicted. Many of them attributed this merely to measurement errors, however, some students speculated that friction was the cause. This led to a discussion in one class of how these results could be used to determine the friction in the experiment.

Conclusion

As a participant in the SEAQL workshop, I was impressed by the enthusiasm the staff had for this data analysis process. Within the first week, I started to understand why, and my experience so far this year, in using SEAQL techniques with my students, has convinced me that it is an indispensable tool for the teaching of science. It emphasizes the importance of data analysis, and provides insights into problems that often did not arise before. SEAQL has given me a new mind set, which I hope carries over to my students.

John Thompson
Valley High School
Spring Hills, Pennsylvania

Workshops Announcement...

Enriching a First Statistics Course Using Prediction Models and Computers A half-day workshop for high school and college teachers of non-calculus-based introductory statistics courses. Concepts from regression and linear models combined with statistical software prepares students to deal with t-tests and analyses of variance under one unifying approach. Where: At annual Joint Statistical Meetings in Anaheim, California. When: Saturday, August 9, 1997 8:30 a.m.–12:30 p.m. Cost: $25 For more information contact Cathy Crocker, ASA Director of Education cathyc@amstat.org (703) 684-1221 ext. 146 or Joe Ward, joeward@tenet.edu (210) 433-6575.

There will also be Quantitative Literacy and AP Statistics workshops at the Joint Statistical Meetings in Anaheim. For more information contact Cathy Crocker at ASA.

The TI-83 Makes the CLT Come to Life!

For years I have been searching for an activity that will lead students to discover (or, at least, believe) the Central Limit Theorem (CLT). I have used digits from random telephone numbers and the ages of randomly selected pennies. I have observed the lima bean sampling of Duane Hinders. None of these activities provided a satisfactory result. In the fall of 1996, I attended a talk by Landy Godbold who used a TI-83 program to created sampling distributions. I prefer not to use programs but Landy's activity sparked an idea which I pursued.

Each of my AP Statistics students was provided with a different value of n. These numbers were varied from 10 to 150. Using the TI-83, the students were to define L1 as "seq(randInt(0,9),x,1,n,1)" replacing the n with the number provided. The quotes are very important, therefore this could not be done as easily on the TI-82.

The definition causes L1 to be filled with n digits selected from the uniform distribution of the digits zero through nine. The student then used the STAT menu to compute the mean of the n digits in List one. The mean was recorded on a piece of paper. When the student returned to L1, a new list of n randomly selected digits replaced the previous list. The new list occurred due to the quotes in the definition of L1. Again the sample mean of the n digits was computed and written down. The two steps were repeated until the student had twenty-five sample means, each for a sample of size n.

Once the twenty-five means were obtained, the student deleted the definition of L1. If this is not done, the data entry which follows becomes a very slow process since a new sequence is executed every time another data item is entered in L2. The twenty-five sample means were entered in L2 and the mean and standard deviation of the twenty-five sample means were computed.

In class, each student provided the sample size, the mean of the sample means and the standard deviation of the sample means for his or her simulations. A scatter plot was obtained with sample size as the X variable and mean of the sample means as the Y variable. The plot (figure 1) appeared to show that as sample size increased, the mean of the distribution was displaying less variability. The class was asked to determine what value the means seemed to be heading toward as the sample size increased. The class obtained the mean of a uniform distribution of the random digits

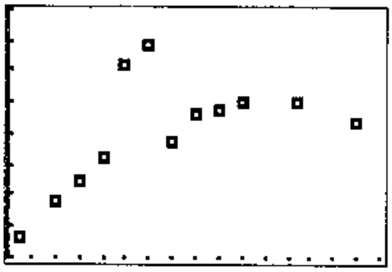

Figure 1

and found it to be 4.5 Y=4.5 was plotted (figure 2) and the students agreed that this seemed to be the value toward which the means of the distributions were approaching as sample size got larger.

The class next looked at a scatter plot of sample size versus the standard deviation of the distributions (figure 3). The plot appeared to be the negative

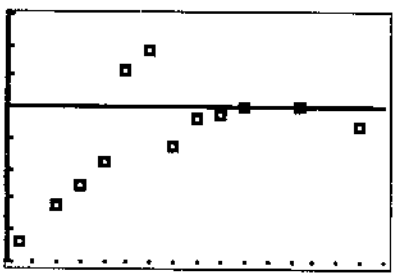

Figure 2

half of a square root function. With some guidance, the class discovered that a constant divided by the square root of X would yield such a graph. The question was, "What was the constant?" It was determined that if the sample size was one, the standard deviation would be the standard deviation of the uniform distribution which

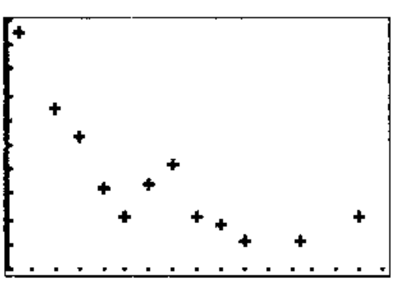

Figure 3

is 2.87. $Y=2.87/\sqrt{x}$ where x was the sample size was placed in the equation editor. The graph of this function (figure 4) fit the scatter plot very well.

Finally the class looked at histograms of the twenty-five means, first for the student who had a sample size of twenty (figure 5) and then for the student with a sample size

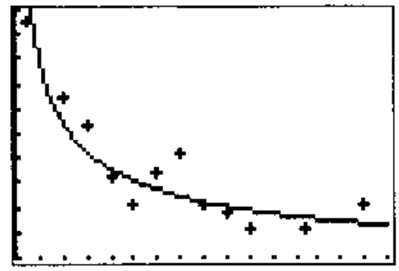

Figure 4

of 150 (figure 6). The sample size for n=20 had no recognizable shape while the distribution of samples of size 150 appeared to have a shape that was approximately normal.

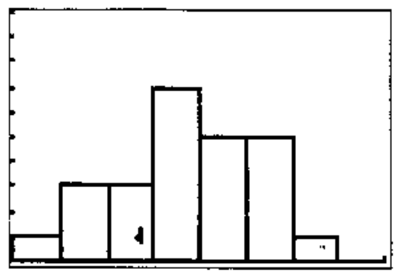

Figure 5

The students had discovered that for a sampling distribution of sample means, as the sample size increases the distribution approaches normality. Also, as the sample size increases, the mean of the distribution of the sample means approaches the population mean and the standard deviation of the distribution approaches the population standard deviation divided by

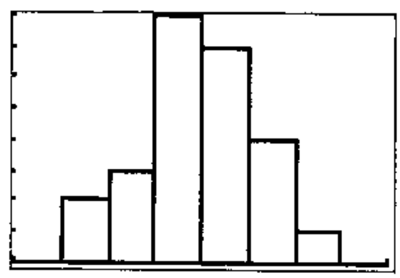

Figure 6

the square root of the sample size. The understanding of inferential statistics is one step closer to realization.

Murray H. Siegel
Marietta High School
Marietta, Georgia

From the Editor

Statistics Teacher Network is on the Web! Thanks to Tom Short of Villanova who will prepare the web version of *STN* and to Mike Meyer and Bruce Trumbo who are the ASA Webmasters. Check it out at: http://www.amstat.org/education/STN/

It was wonderful meeting many of you "E-mail personalities" in person at the NCTM convention in Minneapolis. There were many interesting statistics talks there. Hopefully, next year, there will be a conference within a conference for AP–Stats as well as other focused sessions in Quantitative Literacy and Data Driven Mathematics.

Also, keep in mind that there are several QL workshops scheduled for this July to be held in the Washington, D.C. area. For information, contact Cathy Crocker at the ASA office—by phone (703-684-1221) or by E-mail (cathyc@amstat.org).

For summer homework, take some time to write an article or a review for future issues—or at least drop me a line indicating what topics you would like to have addressed in *STN*. I'm also interested in hearing about your favorite stats Web sites.

Keep in touch. Have a productive summer. See you in the fall!

Keep Us Informed...

The Statistics Teacher Network is a newsletter published three times a year by the American Statistical Association—National Council of Teachers of Mathematics Joint Committee on the Curriculum in Statistics and Probability.

We need your letters, announcements, articles, and information about what is happening in statistics education! Please send hard copy, and, if possible, a disk written in standard ASCII text to the editor:

Jerry Moreno, Dept. of Mathematics, John Carroll University, University Heights, OH 44118

or **moreno@jcvaxa.jcu.edu**

or **fax: (216) 397-3033**

Production Editor: *Kristen Fernekes, American Statistical Association*

To be added to the mailing list or make an address change, please send your name and address to: *Statistics Teacher Network*, c/o American Statistical Association, 1429 Duke St., Alexandria, VA 22314-3415; (703) 684-1221; fax: (703) 684-2036; E-mail: chris@amstat.org

Printed in the U.S.A

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

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

Let us know your Zip +4!