A Tasteful Investigation

of

Elementary Concepts in the Study of Data

Hands-on Workshop (W14)

Concetta M. Duval

Animal crackers, and cocoa to drink,
That is the nicest of suppers, I think.
When I'm grown up and can have what I please,
I think I shall always insist upon [M&M's and] these.
Christopher Morley[1]

Abstract
Animal crackers and M&Ms are fun to look at and delicious to eat. They also provide tasty objects that students can use to learn some elementary statistics. During this workshop, participants will conduct two experiments designed to be used with students at various grade levels depending on their mathematical experience. The statistical questions to be answered require the application of many mathematical skills. And technology, in the form of computer software, provides multiple (and accurate) ways to display the information gathered during an experiment. These displays promote dialogues among teachers and students about mathematics and specifically, statistics.

Introduction
M&M's and BARNUM'S ANIMALS Crackers are not only delicious treats, but provide interesting and delicious contexts in which students can practice their arithmetic, geometry, and algebra skills to learn about statistics. In a world of ever increasing data and information, knowing something about statistics and how to analyze data has become almost a survival skill. Learning statistical skills now begins in elementary school. Starting in the very early grades, students are now expected to conduct surveys, display the results on various types of tables and graphs, and most importantly, decide what the data reveal.

A recent article by mathematicians Allan Rossman and Beth Chance offer ten recommendations for the study of statistics at the college level that focus on "…student investigation and discovery of inferential reasoning, proper interpretation and cautions use of results, and effective communication of findings." [6]

Here, in abbreviated form, are five recommendations of their "top 10" recommendations that seem most directly applicable to younger age students as well:

1. Have students perform physical simulations to discover basic ideas of inference.
2. Encourage students to use technology to explore properties of inference procedures.
3. Always consider issues of data collection
4. Always examine visual displays of the data.
5. Insist on complete presentation and interpretation of results in the context of the data.

With these recommendations as background, then, the participants in this workshop will pose questions related to the delectable treats, M&M's and BARNUM'S ANIMALS Crackers. Then, they will collect and record data, display their results using paper and pencil and computer software, analyze the displays, and answer initial questions based on their findings.

[1] As quoted in [7], with apologies to Mr. Morley and Professor Sommers for the insertion.

M&M's

While grading papers, a Yale researcher, Ronald D. Fricker, observed that there seemed to be far fewer blue M&M's than any of the other six colors in a bowl of plain M&M's. Because he was a statistician (and for the moment, tired of grading papers), he launched a small experiment in which he counted the distribution of colors in the candies and in one 1-pound bag of plain M&M's, found the following distribution of colors: [1]

Brown	35.0%
Yellow	26.7%
Red	15.6%
Orange	8.1%
Green	7.1%
Blue	7.5%

Curious about these results, Fricker dug a little more deeply and uncovered these statistics reported in an article in the *Austin American-Statesman* shortly after blue M&M's replaced tan M&M's in 1995.[2]

Brown	20%
Yellow	20%
Red	20%
Orange	10%
Green	10%
Blue	20%

Clearly, something was amiss, and as Fricker stated, "I could come to only one of two conclusions: either I had uncovered a large corporate conspiracy designed to dupe an unsuspecting public out of blue 'M&M's' or the newspaper was wrong."

So, was the newspaper wrong? Or are we being duped? If so, what does the company say is the distribution of colors of the candies?[3]

Mindful of the recommendations of Rossman and Chance, pairs of participants will conduct a hands-on simulation, collect data, and use a software program called *Graphers* published by Sunburst Communications. As each pair displays its data, they will experience what Rossman and Chance observed: "Students observe first-hand that the outcomes of a statistic vary from sample to sample …. They also notice that while these values differ, a predictable pattern emerges." [4]

Once each group has completed collecting, displaying, and analyzing its own data, the data will be pooled and analyzed as a whole using a second software program—Microsoft's® *Excel*.

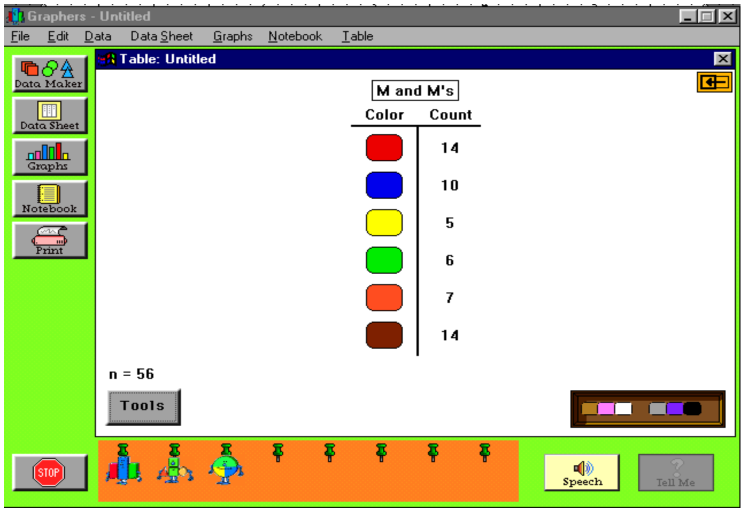
The statistics we will consider are those that students in elementary and secondary students are expected to know: the mean, median, and the mode. The displays to be used are frequency tables, bar graphs, and circle graphs. Since *Graphers* was designed specifically for young students, it will be relatively easy to learn.

Examples of the kinds of displays participants will generate can be seen in the following graphs. Another mathematician, Ivars Peterson, set about repeating Fricker's experiment and came up with his own set of data after conducting several taste tests. [5] Here is a frequency table of one of Peterson's data set done in *Graphers*.

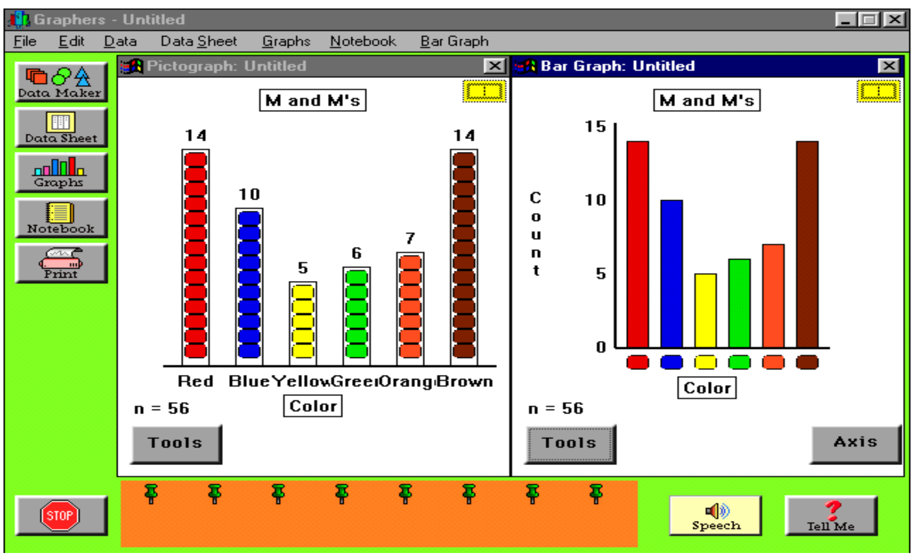
---

[2] According to Fricker, Mars surveyed more than 10,000,000 people to see which color they would prefer over the then existing tan color. Blue won over purple, pink, and no change by 54%, 32%, 10% and 4% respectively.

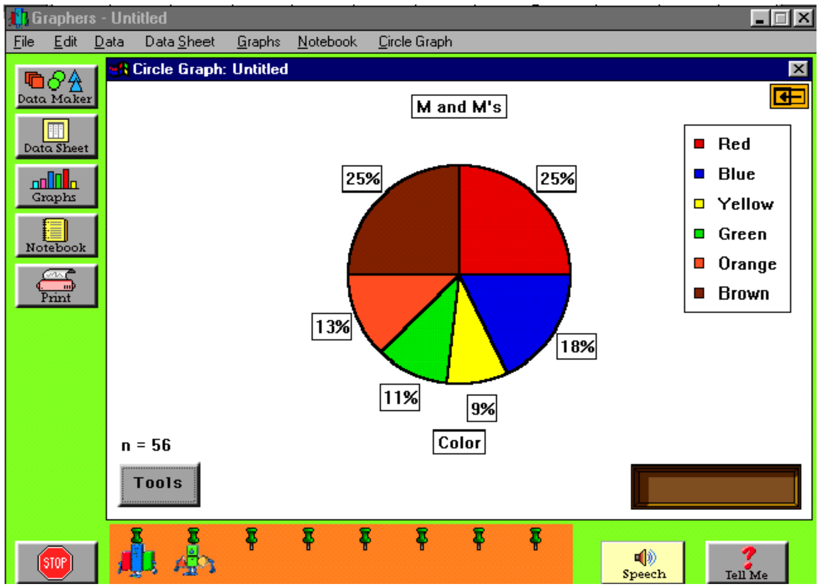
[3] To be revealed at the conclusion of the experiment



A pictograph and bar graph are two other ways to pictorially represent simple frequencies. Using *Graphers,* participants will be able to see the similarities and differences between these two types of graphs by displaying them side-by-side. In this way, they can see that a pictograph is a one-dimensional representation of the count of each color along one axis. A bar graph is a two-dimensional representation of the frequency of each color, requiring two axes.



A fourth type of display available in *Graphers* is a circle graph. Here, again, is an example using Peterson's data. Labels for the sectors can be simple counts, ratios, or percentages, as shown here. Creating circles graphs can be more difficult for younger students because they must know how to use proportions to determine the number of degrees in the central angle of each sector. Thus, the advantage of using a program such as this to draw a circle graph lets students focus on what it means rather than how it's constructed.



As we work through displaying these different types of graphs, participants are to discuss the relationships among the different numeric representations and use them in instruction. For example, using this frequency table of Peterson's data, participants (and eventually their students) can create the fractions, and determine the corresponding percentages.

	n	n/N	%
Brown	14	14/56	25.0
Yellow	5	5/56	8.9
Red	14	14/56	25.0
Orange	7	7/56	12.5
Green	6	6/56	10.7
Blue	10	10/56	17.9
Total	56	56/56	100

After each pair has analyzed the contents of a bag of M&M's, the data will be pooled and analyzed as a whole. A hypothetical aggregate result might look like the numbers in this table. The *Ideal distribution* represents the numbers that would be expected given the percentages reported by Mars.

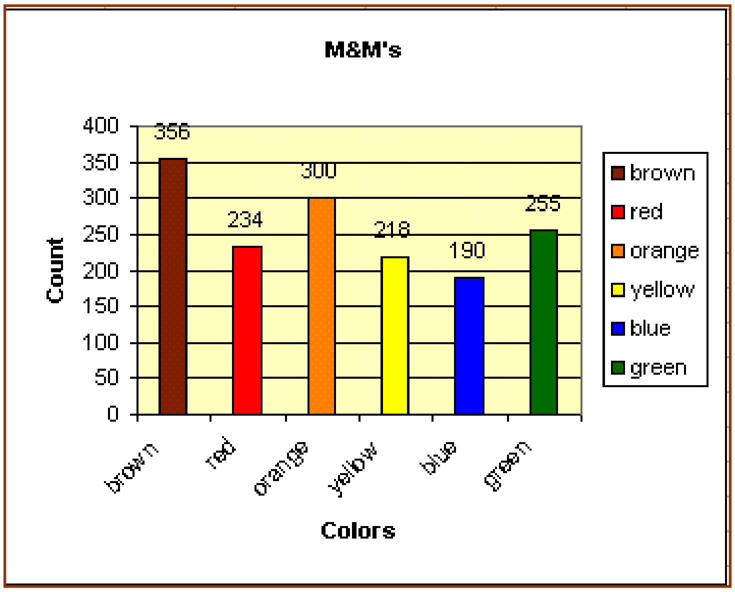
Colors	Number	Ideal distribution
brown	356	466
red	234	311
orange	300	155
yellow	218	311
blue	190	155
green	255	155

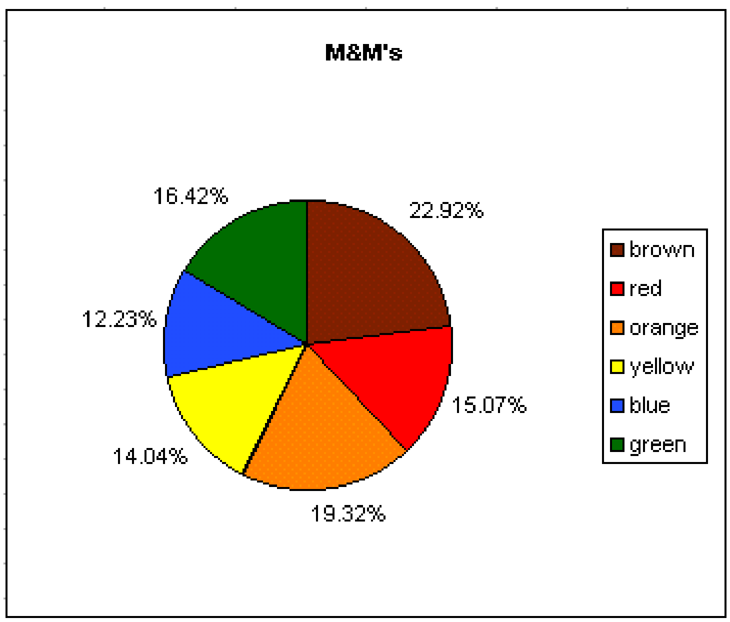
As shown below, the total of candies in this hypothetical sample, and the mean and median of the data can also be calculated in *Excel*. The mean and median for each color M&M can similarly be determined.

Total	**1553**
Mean	**258.8**
Median	**244.5**

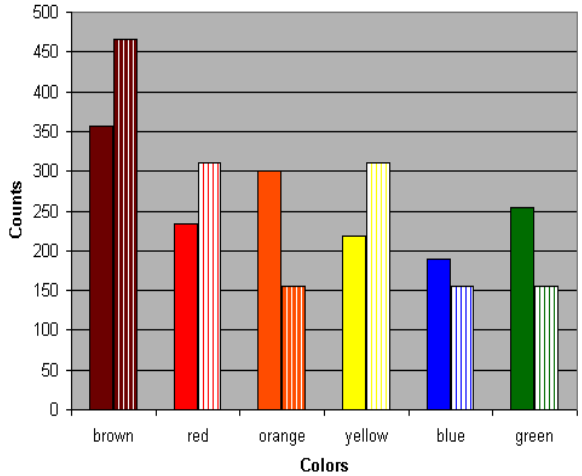
Of course, interpreting these statistics in terms of our experiment becomes crucial. Collecting and displaying data are only a means to an end. The end is interpreting the data to see what they tell us.

Here are the bar graph and circle graph created in *Excel* for these hypothetical results.





To compare the results of our experiment to results we could expect using the percentages of each color as reported by the Mars Company, we could also use *Excel*. This type of display easily gives us a way to establish whether the results reported by the *Austin American Statesman* were correct. (They weren't.) Here is a representation of the data using hypothetical results. The solid bars represent the experimental results for each color, and the striped bars represent the ideal results.



Experiment 2
Nabisco's BARNUM'S ANIMAL Crackers provide a similar context for student investigation. According to Nabisco these treats date from 1902 and were initially marketed as Christmas tree ornaments.[4] Thus, the familiar string. However, unlike M&M's, we don't know at the outset of a statistical experiment what animals we may have in any one box. Today, there are 17 animals: tiger, cougar, camel, rhinoceros, kangaroo, hippo, bison, lion, hyena, zebra, elephant, sheep, bear, gorilla, monkey, seal, and giraffe. The bear comes in two poses; standing and on all fours.

Intrigued by the fact that there are about 20 crackers per box and 17 (or 18 depending on whether the two types of bear poses are taken to be different), Professor Sommers, an economics professor decided to see if there is an average of one animal shape per box. [7] He conducted an experiment involving 109 boxes, finding that there was an average of 11.6 different shapes and 19 crackers per box.

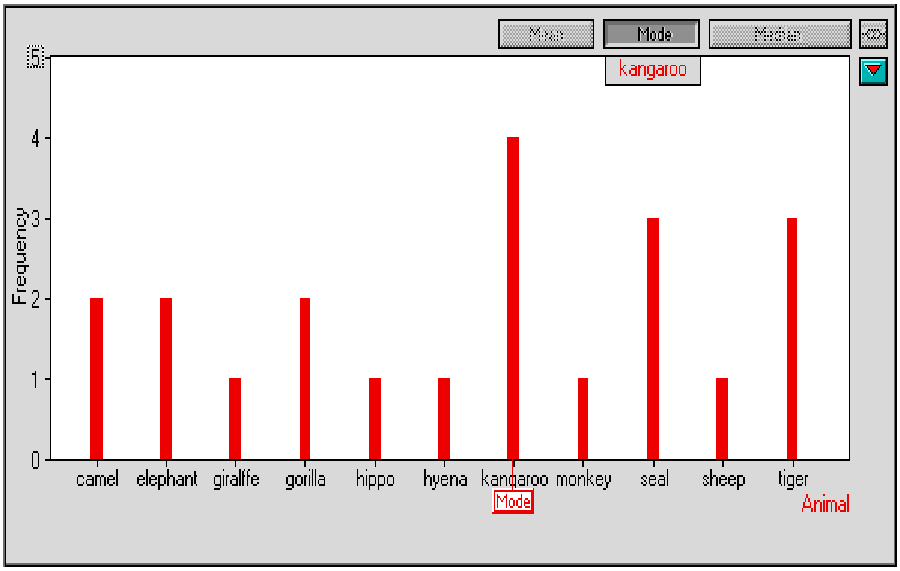
In this experiment, then, participants will identify the number of animal species in each box and as in the M&M's experiment, pool their data to see what the distribution of animals appears to be in a large sample. A statistical program that can be used to graph these data is *Stats!,* part of the *Tangible Math Series* published by Riverdeep Interactive[5]. In this program the data to be collected are entered individually into a spreadsheet. Then, the data can be displayed in a frequency table, a bar graph, and a circle graph.

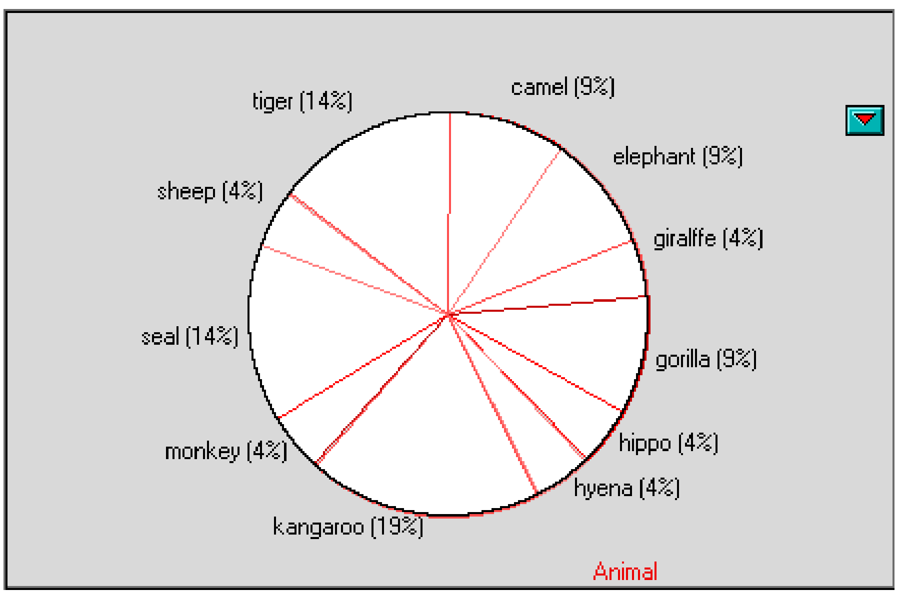
For example, the results of one experiment are listed in this frequency table in *Stats!*.

Animal	n	n/N	%
camel	2	2/21	9%
elephant	2	2/21	9%
giralffe	1	1/21	4%
gorilla	2	2/21	9%
hippo	1	1/21	4%
hyena	1	1/21	4%
kangaroo	4	4/21	19%
monkey	1	1/21	4%
seal	3	3/21	14%
sheep	1	1/21	4%
tiger	3	3/21	14%
Total	21	1	100%

[4] http://www.easycheese.com/school/funfacts/barnums.html
[5] www.riverdeep.net

Here are the bar graph and circle graph created in *Stats!*. The bar graph also reports the mode of the data.

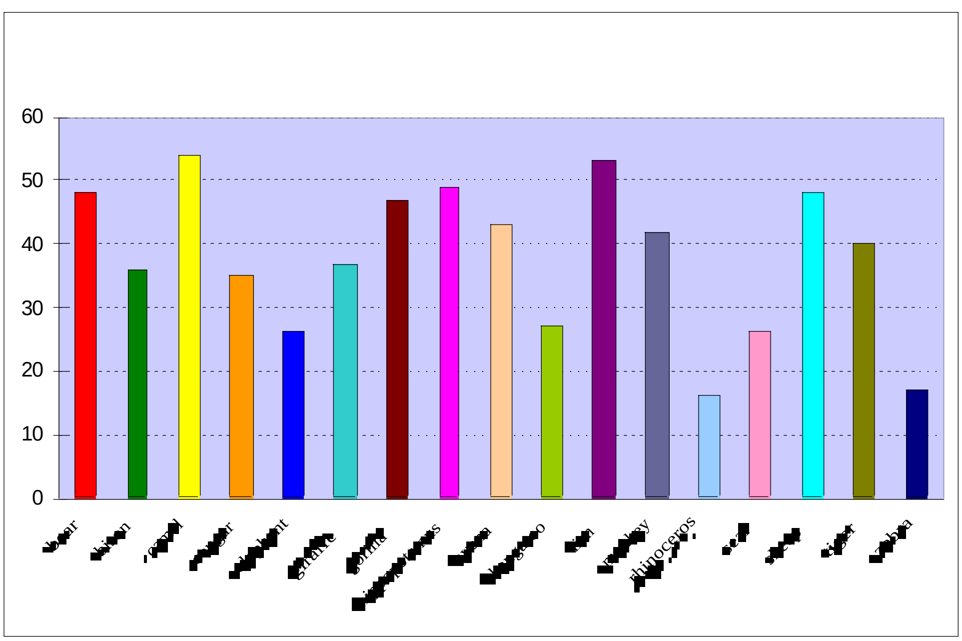




Because the aggregated data will include all of the animals, *Excel* is an excellent way to display the total count for all the participants. Using hypothetical data, participants enter the names of all 18 (or 19) types of crackers into one column and the total number of each type in a second column. The mean number of species per box can be calculated automatically in the third column by entering the proper formula.

Animals	Number	No./box
bear	48	1.9
bison	36	1.4
camel	54	2.1
cougar	35	1.4
elephant	26	1.0
giraffe	37	1.4
gorilla	47	1.8
hippopotamus	49	1.9
hyena	43	1.7
kangaroo	27	1.0
lion	53	2.1
monkey	42	1.6
rhinoceros	16	0.6
seal	26	1.0
sheep	48	1.9
tiger	40	1.6
zebra	17	0.7
total	**644**	

These data can then be displayed in a bar graph such as this.



This particular experiment offers other opportunities for exploration. One such investigation suggested by Professor Sommers is to sort the animals into two categories: predators (carnivores) and preys (herbivores). First, of course, students might have to learn more about each animal. Then, knowing the difference between the two types of animals, students could see if there was more of one type of animal than another in a box. Are the predators devouring the prey?

There are no doubt other questions that participants and eventually their students might wish to investigate. For example, for advanced students of statistics, students could perform a chi-square analysis as done by Professor Sommers. Follow-up discussions at the end of each experiment may indeed include such questions that require additional investigation.

Conclusion

These two experiments are ways to incorporate experimentation into mathematical classrooms at the upper elementary, middle, and even high-school levels. The statistical questions to be answered require application of skills in arithmetic, geometry, and algebra. Various computer software programs provide the means to display the results of each experiment and to use these displays to analyze and communicate what the data mean. As Rossman and Chance conclude in their article, “Statistical education reform emphasizes active learning on the part of students, conceptual understanding of fundamental statistical ideas, use of engaging applications involving genuine data, and development of student communication skills.”

References

1. Fricker, Ronald D. Jr., *The Mysterious Case of the Blue 'M&M's'®,* Yale University, Department of Statistics, June 8, 1996.
2. *Graphers*, Sunburst Communications, Inc., Pleasantville, NY, 1996.
3. Microsoft® *Excel* 97, Microsoft Corporation, Bellingham, WA
4. Nabisco, http://www.easycheese.com/school/funfacts/barnums.html
5. Ivars Peterson's **MathLand**, http://www.maa.org/mathland/mathland_3_10.html
6. Rossman, Allan and Chance, Beth, Teaching the Reasoning of Statistical Inference: A 'Top Ten' List, *The College Mathematics Journal*, Vol. 30, No. 4, September 1999, 297 – 304.
7. Sommers, Paul, Students' *t* and Crackers. *The College Mathematics Journal*, Vol. 30, No. 1, January 7 1999, 32 – 34.
8. *Stats!, Tangible MATH™ 3.04 Series*, Riverdeep Interactive, Cambridge, MA, 1998.