MATHEMATICS IN A STATISTICAL CONTEXT?

Linda Gattuso
Université du Québec à Montreal, Canada
gattuso.linda@uqam.ca

Statistics is often taught as part of the mathematical curriculum in schools by mathematics teachers who often do not feel prepared to include it in their practice. They feel that they are stealing time that would be better used if dedicated to mathematics. We believe, however, that it is possible to develop a collaboration between the teaching of mathematics and the teaching of statistics. Using statistics as a realistic context to work on certain mathematical concepts can provide significant motivation for students. To accomplish this, it is essential to highlight the mathematical concepts underlying statistical concepts in order to link them in creating classroom activities that are also useful for teacher training. In this paper, we will try to illustrate this idea starting with examples gathered in school mathematics and in basic statistics.

INTRODUCTION

Statistics is necessary for today's citizen, and without this knowledge it is difficult to participate in social and political debates (Konold, 2003). However, many teachers are not comfortable with data analysis and lack confidence in teaching it (Russell, 1990a). Teachers missing specific training get their information directly from school textbooks without the possibility of detecting the errors present, and are unable to recognize and respond to students' conceptions and misconceptions (Russell, 1990a). Interviews of mathematics teachers reveal "a feeling of insecurity due to a lack of preparation in the teaching of statistics" (Gattuso & Pannonne, 2002). Furthermore, teachers not aware of the richness of the statistical content think that descriptive statistics is easy and not very interesting. Consequently they relegate statistics to the end of the school year or leave it out completely because of lack of time (Aksu, 1990).

Statistical and mathematical reasoning are different and it is important to distinguish between them. In statistics, context is essential and inherent to the analysis, whereas mathematics may be context independent, concentrating on abstract numbers. However, today's schools increasingly present mathematics as a tool to cope with everyday problems. In reality, numbers are found in comparisons, tables, orders of magnitude, rounding, estimating, prices and other numerical messages. Children must "develop [their] mental images of numbers parallel to their acquisition of counting and calculation skills" (Dunkels, 1986, p. 61). Statistics that describe the world around us contain mathematics. Real contexts used in statistics can support and contextualize mathematics topics and also help to motivate students.

Mathematics often asks for one correct solution where statistics ends with various possibilities and uncertainty. Nowadays, school mathematics has adopted new approaches in which problem solving and open-ended problems call for various creative solutions. Statistics favors a less restrained method and helps students deal with uncertainty.

While mathematics involves deductive reasoning, statistics calls for inductive reasoning; however, before getting to a formal proof, even mathematicians deal with examples and induction. This means that in mathematics, observation alone is not sufficient, while in statistics everything starts with observed results. So where mathematics can help to structure and be accurate, statistics can encourage diversity and creativity.

In this paper, we will try to call attention to the fact that the teaching of statistics can assist the learning of mathematics because it includes many mathematical concepts in a realistic and motivating context. It is necessary to match mathematical concepts to their use in statistics so that one supports the development of the other (Dunkels 1990). One of the reasons for this is to address teachers' insecurities by assuring them that while doing statistics they are building mathematical skills. Since statistics is taught mostly by mathematics teachers, it is important to show them that allocating time for statistics does not in fact undermine the importance of mathematics but may actually reinforce its learning. An analysis of selected school textbooks and descriptions of experiments revealed that the statistical concepts introduced contained most

© ICMI/IASE 2008. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.*

of the mathematics included in the elementary curriculum and part of the high-school curriculum.

DATA ANALYSIS AND REPRESENTATIONS

Data analysis usually begins with descriptive statistics, often with small surveys, simple activities that can be beneficial even to preschoolers (Dunkels, 1991). Young children from 4 to 8 years old usually fix their attention on *an* individual and not *the* individual (Russell, 1990b), but collecting and organizing data methodically brings them to acknowledge the existence of others and the existing similarities or dissimilarities. This serves as an initiation to the concept of variability (in mathematics as in statistics) as it is constantly active in data handling activities.

The construction of graphical representations is also useful. In a first step, it is important to let the children create their own representations, and from there bring them to see the elements that are essential and the ones that are missing. Progressively, traditional graphics will be presented, leading to the study of functions and their graphs.

Once the data are organized and displayed in a graph or a table, it is time to examine "what the display tells us?" (Dunkels, 1991, p 131). Some questions will require counting or ordering and will bring about the distinction between cardinal and ordinal numbers without the introduction of important terms such as "less than, at least, more than" and so on. Other questions may require addition, subtraction and multiplication. These might be done by counting the dots (Figure 1) with a one-to-one correspondence before applying (with older children) the adequate operation.

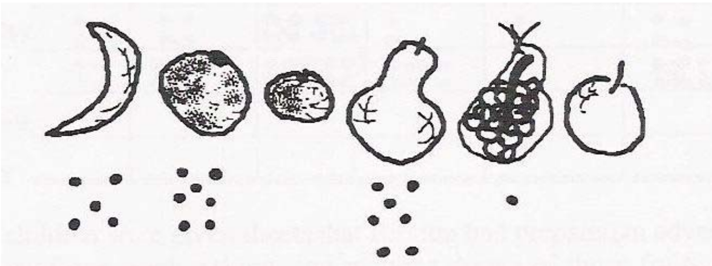

Figure 1. The fruit diagram (Dunkels, 1991)

Numbers will eventually replace the dots and words will take the place of the drawings representing each category of data. Simple pictograms will prepare the way for organized tables where the categories will be replaced by numerical data and by intervals. The concept of a number as a measure will appear, for example for height or weight. Because the data are quantitative, order becomes essential. Contexts will become more complicated and data will, in the end, draw on bigger numbers, real, rational, and decimal numbers, requiring the related arithmetical operations.

More or less simple stem-and-leaf plots require understanding of the sense of a number second to its position, units, tens, etc.. It is also necessary to order the data, and if the numbers are big, estimation may be needed (Dunkels, 1986, 1991).

Constructing a pie chart provides an occasion to work with fractions as part of a whole, using multiplication to find the portion of 360° representing a certain characteristic. This is an opportunity to work on proportional reasoning, equivalent fractions and corresponding percentages and vice-versa.

A bar chart requires dealing with scales and frequencies that are proportional to the length of the bars. Constructing relative frequency tables will also apply the same types of reasoning. Time is a continuous variable and introduces class intervals, open or closed. Construction and understanding of histograms also demands proportional reasoning. The important point is the proportional link between a frequency expressed as an integer, a fraction, or a percentage, and the area, which is a two-dimensional measure. This is a crucial step for the construction of the concept of distribution function. Histogram drawing also requires the minimum and maximum of a sequence of numbers, comparison of numbers, counting frequencies in intervals, obtaining relative frequencies or proportions, etc.

Box-plots helpful for comparisons will accentuate the work on the number line, the concepts of half and quarter and also of 25%, 50%, 75% and range, and underline the distinction of the role of numbers: data, frequency, relative frequency. Taking advantage of multiple pictorial forms to analyze the data is important, since graphical methods seem to stimulate thinking, contrary to formulas that often rely on automatism (Vännman, 1990).

If students question whether their choice of pets is similar or different from that of children living in other countries, they will find the data on the Internet, for this and numerous others variables, since a wide choice of databases are freely available for educational institutions. To compare data coming from different sources, pupils will have to transform the frequencies into percentages and look into the relative modification of the percentage for different totals, leading to a more profound insight introduced later. For instance, for a group of 20, two more is 10% more, but if the group size is two thousand, what difference in the size will generate a variation of 10%?

An introduction to the notion of dependence between two categorical variables requires double entry tables and the transformation of the data into percentages by columns or rows for comparison. Cross tabulation gives a meaning to proportions, equal and near equal proportions. Proportionality (including percentages) is a fundamental mathematical concept that brings its share of difficulties (René de Cotret, 1991; Pézard, 1986), and statistical tables and graphs offer a good opportunity to put proportionality into practice.

CENTRAL TENDENCY AND VARIABILITY MEASURES

Good questioning can activate new reasoning and new concepts, for example: “Half of the mothers are older then….?” or “Is it possible to find the value that divides the data into two equal groups?” Separating each new group in two, thus separating the original group into four parts, then makes it possible not only to identify the median and the quartiles but again to work on rational numbers–half and quarter–and on ordinal numbers.

Position measures can be introduced quite early. The computing algorithm for the mean uses addition and division in the sense of “equal sharing” and the resulting quotient can also be seen as a ratio. The concept of spread, as in the range, calls for a difference and the notions of greater and smaller. These concepts can be used not only to exercise addition and division but, more importantly, to confer meaning to arithmetical operations. For grouped data, the introduction of central tendency and dispersion measures can lead to a discussion of estimation, since results obtained with the raw data may differ from the ones found using the middle value of the interval. This could be introduced later.

Using the computer to find position measures will also leave more room for interpretation and analysis. In doing so, not only does the student familiarize himself with the use of these “machines”, but he also needs to develop a critical sense about multiple productions. In addition, technology leaves room for the development of imagination and creativity, questioning, analyzing and interpreting. The student can then ask questions such as “Does this graph make sense?” (Huff, 1954), or “the chi-square value is…”, or “What does it mean?”

BIVARIATE RELATIONSHIPS

The study of correlation and regression exploit the concepts of variability and functions, beginning with scatter plots that use the plane, axes and coordinates. From there, it is possible to trace a line estimating the relation between the points and look for the equation of this line starting with various information: two points, or the y-intercept and the slope, and so on. Later, the students will look for the line (or eventually, the curve) that best translates the relation between the two variables via the concept of distance between two points. The line graphs representing the relationship between two variables can be used to observe trends and to make predictions, as another way to interpret numbers, and eventually to build models.

Use of a spreadsheet (like Excel) makes it possible to use trial and error to find which straight line would be the best estimation of the relationship between the two variables involved. After drawing a straight line passing as near as possible to all the points of the scatter plot, a student can find the equation using two points, or using the slope and one point. For each point, the difference between the ordinate of each point, the observed values (*y*), and the “predicted” ordinate, values ”predicted” by the drawn line (Y)–given by the equation of the estimated straight line–is calculated. Based on the points and the equation of the straight line, using the spreadsheet capacities, Table 1 can be quickly filled in.

Table 1. Spreadsheet table of y-*Y*

x	y	$Y=mx+b$	y-Y
:	:	:	:
		sum=	

With an extreme example (where the sum of y-$Y = 0$) the teacher can challenge the fact that the sum of the differences is the best way to treat this variation (because positive and negative differences cancel each-other out) and lead to the exploration of the notion of distance between two points (absolute value) and of the squares of the differences between the y and the *Y*, and of their respective sum. The teacher may seize the occasion to explore absolute values, absolute value function, and quadratic function and their minima.

“To find the straight line that is closest to all points, should we use the sum of the distances or the sum of the squares of the differences?” The common question about the choice of the sum of the squared differences over the absolute value of differences can be answered by studying in parallel a graphic representation and the corresponding algebraic manipulations.

The sum of two quadratic functions:

$$
\begin{gathered}
(y_1-y)^2+(y_2-y)^2 = \\
(y_1^2-2y_1y+y^2)+(y_2^2-2y_2y+y^2)= \\
(y_1^2 y_2^2)+2(y_1+y_2)y+2y^2 = \\
c+by+ay^2
\end{gathered}
$$

Another quadratic function!

A simple graphic tracer when asked for the graph of the sum of two quadratic functions (Figure 2) will produce a quadratic function whose minimal value is easily found.

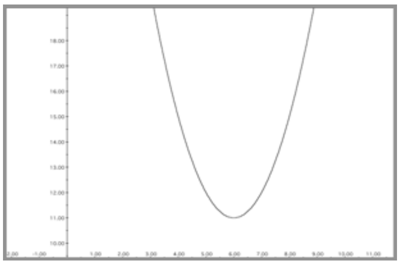

Figure 2. Sum of two quadratic functions

Whereas the sum of two absolute value functions (Figure 3) is:

$$|y_1-y| + |y_2-y| = ?$$

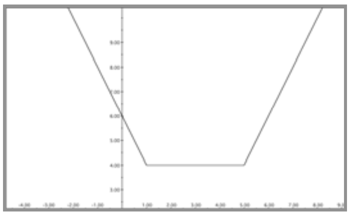

Figure 3. Sum of two absolute value functions

Observation of the graph of the sum of two absolute value functions shows that the new function obtained does not have one minimum value but an interval of minimum values. Secondary school students are able to reason why the sum of the squares of the differences was preferred without going into formal proof using calculus. It is now possible to add a column of squares: (y-$Y)^2$ and to sum them up.

Table 2. Spreadsheet table with the (y-$Y)^2$ column

x	*y*	*Y=mx+b*	*y-Y*	| *y-Y* |	$(y\text{-}Y)^2$
:	:	:		:	:
:	:	:		:	:
				SUM	SUM

The computer allows the students to “move” the line by varying the value of *m* and of *b* in order to find the minimal value for the sum of the squared differences and, by so doing, to find a line that gives a better estimate of the relationship between the two variables. This relationship provides the occasion for discussing the difference between causality and correlation, a distinction important for statistical analysis. The line can be used in prediction exercises and can also illustrate the difference between interpolation and extrapolation. In fact, this line is also an example of model construction of a phenomena.

DISCUSSION

The idea we defend in this paper is not so much to teach mathematics but to be aware of all the mathematics included in the practice of statistics. Statistics offers a context that favors an understanding of mathematical concepts countering a formalist and void presentation, and brings in a non-deterministic method that favors a more creative approach to mathematics.

Even when a curriculum includes statistics, the time allocated to the subject is mainly controlled by the teachers. However, because they also often teach mathematics, it is important that they realize the positive contribution that statistics can make. Not only does statistics bring in a meaningful context and a creative approach, which can stimulate the students’ interest in “numbers” and mathematics, for which there is a great need, but it also utilizes many mathematical concepts taught in schools.

Moreover, with a basis in data, students not only deepen their understanding of mathematics, they also encounter a discipline that occurs more and more in their everyday environment, familiarizing them with a point of view where answers are not unique and straightforward but where arguments are based on quantitative results, as all science should be.

If we highlighted some of the mathematics found while doing statistics, it is not to suggest a more formalist approach to statistical knowledge but to convince mathematics teachers that time devoted to statistics can enhance the learning of mathematics. Teachers also need to be able to construct a conceptual analysis of the statistical concepts taught in order to identify and address the difficulties underlying them; for example, proportionality in the construction of pie charts. On the other side, being conscious of the mathematics involved can help to clearly distinguish the part that addresses specific statistical thinking, and thereby construct learning activities that necessitate the integration and activation of statistical reasoning.

From elementary to high school, the study of statistics is an excellent opportunity to use calculators (graphic or not), computers and the Internet. Processing large quantities of data, which shows the usefulness of statistics, also confers particular importance to the integration of technological tools in mathematics. Thus, the student is dispensed from repetitive operations that are not very stimulating and that could not be processed otherwise in a reasonable amount of time.

“Statistics should not be taught as a separate unit, but should be introduced whenever appropriate to illustrate and expand upon standard concepts (such as measurement) and to form interdisciplinary links for students" (Burrill, 1990, p 222). Vännman says, "…statistics can be involved in the mathematics course without necessarily "throwing out" important mathematics areas. Instead, mathematics is strengthened by discussions around statistical matters, and using statistics can introduce reality into mathematics courses" (1990, p 120). Nevertheless, being conscious of the mathematics included in working with statistics is not enough to develop statistical reasoning and learning in the students. Therefore we must continue to promote innovative teacher training experiences where the principal aim is to develop statistical reasoning.

REFERENCES

Aksu, M. (1990). Problem areas related to statistics in training teachers of mathematics in Turkey. In A. Hawkins (Ed.), *Training Teachers to Teach Statistics Proceedings of the International Statistical Institute Round Table Conference, Budapest* (pp. 127-137). Voorburg, Netherlands: International Statistical Institute.

Burrill, G. (1990). Quantitative literacy: leadership training for master teachers. In A. Hawkins (Ed.), *Training Teachers to Teach Statistics Proceedings of the International Statistical Institute Round Table Conference* (pp. 219-227). Voorburg, Netherlands: International Statistical Institute.

Dunkels, A. (1986). EDA in the primary classroom. Graphing and concept formation combined. In R. Davidson, & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 61-66). Victoria, B. C.: International Statistical Institute.

Dunkels, A. (1990). Examples from the in-service classroom (age group 7-12). In A. Hawkins (Ed.), *Training Teachers to Teach Statistics Proceedings of the International Statistical Institute Round Table Conference, Budapest* (pp. 102-109). Voorburg: Netherlands: International Statistical Institute.

Dunkels, A. (1991). Interplay of the number concept and statistics using attitudes and techniques of EDA. In D. Vere-Jones, S. Carlyle & B. P. Dawkins (Eds.), *Proceedings of the Third International Conference on Teaching Statistics* (vol.1, pp. 129-139). Voorburg, Netherlands: International Statistical Institute.

Gattuso, L., & Pannone, M. A. (2002). Teacher's training in a statistic teaching experimentation. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics* (pp. 685-692). Durban, South Africa: International Statistical Institute.

Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton.

Konold, C., & Higgins, L.T. (2003). Reasoning about data. In J. Kilpatrick, W.G. Martin & D. Schiffer (Eds.), *A research companion to principles and standards for school mathematics* (pp. 192-215). Reston, VA: NCTM.

Pézard, M. (1986). *À propos de l'enseignement de la proportionnalité* (In relation to the teaching of proportionality). *Les cahiers de didactique 20*. Paris: IREM de Paris7.

René de Cotret, S. (1991). *Étude de l'influence des variables, indice de proportionnalité du theme et nombre de couples de données sur la reconnaissance, le traitement et la compréhension de problèmes de proportionnalité chez des élèves de 13-14 ans* (Influence of variables, proportionality index and number of data pairs on recognising the treatment and understanding of proportinality problems by 13-14 year old students). Unpublished Ph.D. Université Joseph Fourier, Grenoble I.

Russell, S., (1990a). Issues in training teachers to teach statistics in the elementary school : a world of uncertainty. In A. Hawkins (Ed.), *Training Teachers to Teach Statistics Proceedings of the International Statistical Institute Round Table Conference* (pp. 59-71). Voorburg, Netherlands: International Statistical Institute.

Russell, S., (1990b). Counting noses and scary things: Children construct their ideas about data. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 158-164). Dunedin, New Zealand: International Statistical Institute

Vännman, K. (1990). Some aspects of statistical graphics for secondary school teachers. In A. Hawkins (Ed.), *Training Teachers to Teach Statistics Proceedings of the International Statistical Institute Round Table Conference* (pp. 110-125). Voorburg, Netherlands: International Statistical Institute.