HIGH SCHOOL TEACHERS' REASONING ABOUT DATA ANALYSIS IN A DYNAMICS STATISTICAL ENVIRONMENT

Santiago Inzunsa and José A. Juárez
Universidad Autónoma de Sinaloa, México
sinzunza@uas.uasnet.mx

In this article results of a study about statistical reasoning that high school teachers develop in a computer environment are presented. The results indicate that despite the abundance of representations provided by the software, teachers focus on the calculation of averages to describe and compare distributions, rather than on the important properties of data such as variability, shape and outliers. Many teachers were able to build interesting graphs reflecting important properties of the data, but cannot use them to support data analysis. Hence, it is necessary to extend the teachers' understanding on data analysis so they can take advantage of the cognitive potential that computer tools to offer.

INTRODUCTION

In the last two decades the importance of approaching the teaching of statistics towards the development of cognitive processes as statistical reasoning and thinking; more than in calculations, procedures and formulas has been recognized; especially since the statistics has been conceptualized as a science of data in context and not as a branch of the mathematics (Moore, 1992), but also for the development of computer technology that has transformed some phases of data analysis into trivial processes (e.g., descriptive measures calculations, graphs construction).

The new approach has implied a profound change in the statistics curriculum to adapt it to the new learning objectives; at the same time it requires a drastic change in the instructional teachers' practices (Garfield, 2002). Although, some research (e.g., Eichler, 2008) reveals that in respect to this last aspect, the change has not been immediate and has generated diverse difficulties to the teachers. Particularly in the case of Mexico, the statistics curriculum in the most universities and the instructional practice of many statistics teachers still makes too much emphasis in the traditional statistics (calculations, procedures, formulas) and gives little importance to the reasoning and statistical thinking (Inzunsa & Juarez, 2007).

Therefore, we have considered investigate about the impact of a sequence of activities with the support of software Fathom in the reasoning that teachers have on data analysis, particularly about the distribution concept, that involves important concepts such as averages, variability and graphics representations. Our main purpose is to assess to what extent the teachers are conscious of the importance of using software in their instructional practice.

THEORETICAL CONSIDERATIONS

Computer technology has been an element that has motivated the change of approach from the traditional teaching of statistics to the teaching focused in the thinking and statistical reasoning, and should become a valuable tool for the teachers. The computer environments provide dynamic and interactive representational tools that go beyond than "amplifying" the computational capabilities; when used appropriately the computers can generate a "cognitive reorganization" in the mind of the users (Pea, 1987).

In particular, in the teaching of probability and statistics, the computer technology has shown a great potential to help the students understand difficult concepts (Ben-Zvi, 2000; Mills 2004; Change y Rossman 2006). In the probability case, one of the most relevant aspects in which the computer may be useful as a pedagogic tool is the simulation of random phenomena; since with the use of simulation, concepts and principles can be explored and understood (for example: probability distributions, random sampling and sampling distributions) that in another way would be more abstract.

On the one hand, in statistics the computer can be useful to perform laborious calculations (for example; calculations of the standard deviation and the correlation coefficient), in the exploring real data problems and in the construction of graphs (Chance et al., 2007). Given their interactivity and dynamism in the representations, the computers promote the formulation of

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

conjectures, explore and analyze data, promote active learning and motivate the interchange of ideas in collaborative environments (Cobb & McClain, 2004; Garfield & BenZvi, 2008).

METHODOLOGY

Six high school teachers that were taking a course about the new approaches of teaching statistics participate in this research. The design of the activities was planned so that the teachers analyzed quantitative variables separately first, and later made an analysis of a qualitative variable versus a quantitative variable with the objective of establishing comparisons between distributions and use concepts as averages, variability, shape and outliers. Curcio (1989) calls the first activity "read the data" and the second activity "read between the data."

The instructions in each activity indicated to the teachers to use all the resources of the software that were necessary to make the complete analysis and respond to certain questions that pretended to capture the type of representations they used to answer. In total three activities were set (backpacks, cholesterol and quality control); but due to space issues, in this work we will only present the results of the first two activities. The questions of each activity were responded by the teachers in the text boxes provided by the software, from which we have taken some fragments that are shown in the next paragraphs. In the final part of the study two teachers were interviewed with the purpose of deepening into the comprehension they developed during the course.

RESULTS AND DISCUSION

Activity 1

Many students present problems of back pains. The doctors consider that these problems are caused by the weight of their backpacks. Sometimes, the form in which they carry their backpack also contributes to the back pains (Konold & Miller, 2005). Analyze and describe the data in the most complete possible way. Use all the resources of the software that you consider convenient for the analysis and answer the following questions:

- What students do you think carry more weight in their backpacks, the ones from lower or higher grades? Justify your arguments.
- Which students do you think carry more weight in their backpacks, women or man? Justify your arguments.
- If doctors recommend that the weight of the backpack of a student should not be more than the 15% of their body weight, what percentage of students take a weight higher than 15% of their body weight in their backpacks? In these students are there more girls or boys? Are there more superior or inferior grade level students?

First question

All teachers answered correctly the question noting that the ones with a higher grade carry more weight in their backpacks, for which they have taken into consideration the average weight of each of the grade levels. In Fathom it is simple to construct a table with two variables (see table 1), which was the principal resource used to answer the question. It was observed that all teachers constructed diverse types of graphs, but their analysis was based particularly on the table. For example, in the figure 1 that was used by Paul, has marks that indicate the average weight of each grade, but it also gives very important information about the distributions shape, the variability and outliers, but it was not considered by the teacher when answering the first question.

> Just one teacher establishes relationships beyond the averages, but in a superficial manner.
> *The average weight of the student's backpacks is as follows: the first grade shows 4.94 pounds, the third grade is 6.52 pounds, the fifth grade is 11.5 pounds and the seventh grade is 17.42 pounds…The students of a greater grade carry a higher weight, those in fifth grade carry 50% more weight in pounds than those in first and third grade approximately (Gisela).*

On the other hand, Ramon is the only teacher that does not use the table (Table 1) and is based only in the graph (Figure 1), which is insufficient to make the analysis; also, instead of using the averages to define the scholar grade that carry a higher weight in their backpacks, he uses maximum data.

> *We found that the higher grades tend to carry backpacks of higher weight. The graph (dots plot) shows how the students of the fifth and seventh grade come to carry backpacks of a higher weight, even up to 39 pounds, on the other hand the students form first and third grade come to carry up to 10 and 11 pounds maximum.*

Peso Mochilas

		Peso_Mochila
Grado	Five	22 11.5
	Seven	19 17.421053
	Tres	21 6.5238095
	Uno	17 4.9411765
Column Summary		79 10.189873

S1 = count ( )
S2 = mean ( )

Table 1. Average weight of backpack by group

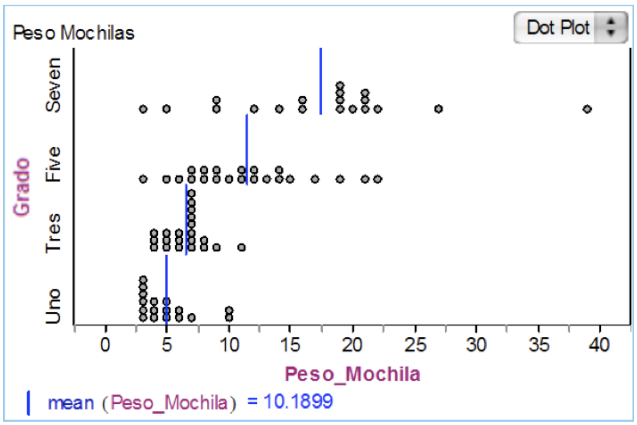

Figure 1. Dot plots with data from each group

Second Question

Again, the teachers answer the question correctly using the averages, mostly by means of a table. In this case we observe that only one of the teachers (Claudia) used the descriptive measurements, without recurring to the graphs, to make the analysis; she constructs table 2 for her explanation; some teachers used more resources of the software to make the analysis than in the last question, such as Claudia's case, who used table 2 for her explanation:

> *The women carry a average weight of 9.92 pounds and the man the average weight is of 10.45 pounds. The half of the women carry a backpack weight of more than 7 pounds [has looked at the median] with a maximum value of 22 pounds and an atypical value of 39 pounds [has looked at the graph]. Half of the men carry a weight higher than 8.5 pounds [has looked at the median] with a maximum of 27 pounds. About the variability it was very similar in both genders with a standard deviation of 1.16 in woman while in man it was 1.02.*

Peso Mochilas

	Genero		Row Summary
	F	M	
Peso_Mochila	39	40	79
	9.9230769	10.462931	10.196421
	3	3	3
	5	5.5	5
	7	8.5	8
	13	15.5	14
	39	27	39

S1 = count ( )
S2 = mean ( )
S3 = min ( )
S4 = Q1 ( )
S5 = median ( )
S6 = Q3 ( )
S7 = max ( )

Table 2. Descriptive measurements by gender

Third question

The difficulty of finding a formula that allows them to answer the question, the teachers decide to construct diverse graphs but choose principally the dots plot as the most adequate graph (Figure 2 and Figure 3), since it allows the counting of data above the recommendation of doctors (15% of body weight) marked by a vertical line.

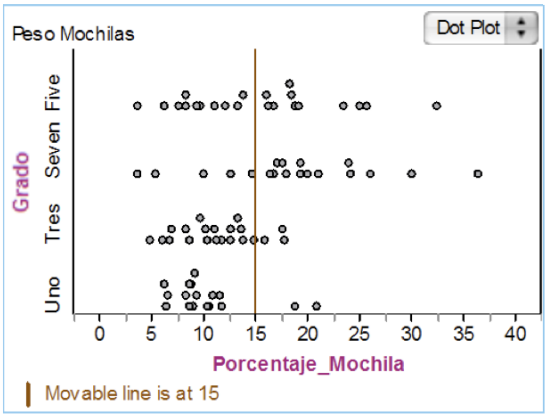

Figure 2. Percentage of weight vs. grade

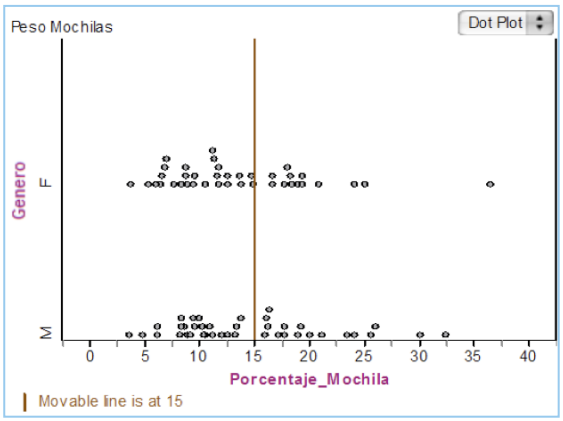

Figure 3. Percentage of weight vs. gender

> *There are more boys (17 of 40) than girls that exceed the recommendation. By grade, there are 14 students of seventh grade who carry more than 15% of their body weight (Claudia).*
>
> *38% of students carry a weight of more than 15% of their body weight in their backpacks, 42.2% of boys exceed the suggested weight and the girls only the 33.33%. The fifth and sixth grades are the ones that mostly exceed the weight by 14% and 18% respectively (Gisela).*

It is important to noting the change of representation that the teachers used in respect to the previous points, where they had principally used a table. Inserting a line that defines the excess limit of weight in a dot graph with both variables of interest is a graphic representation that the software permits to construct and that it was chosen adequately by the students of a group of possible graph representations.

Activity 2

The high levels of cholesterol are a risk factor for heart disease. The data that is shown next corresponds to the cholesterol levels of 24 patients before and after taking a diet based of the consuming of vegetables during the month. The cholesterol levels are measured in milligrams of cholesterol by deciliter of blood (mg/dL). Analyze and describe the data in the most complete way possible. Use all the software resources to respond the following questions:

- Based on the data analysis, do you consider that the vegetarian diet had an effect in the reduction of cholesterol levels? Justify your answer.
- If doctors recommend a cholesterol level lower than 200, what percentage of patients is above such a value before and after the diet?

An adequate table to answer the second question was constructed by the most teachers. Following is a table (Table 3) and graph (Figure 4) constructed by Refugio and Claudia respectively:

> *We can observe that the levels of cholesterol average before the program were of 187.79 mg/dl, while after the program the cholesterol average levels decreased to 168.25 mg/dl [she is looking at the mean]. After the program half of the patients had a cholesterol level*

under 165 mg/dl while before the program the cholesterol levels were 179 mg/dl [she is looking at the median](Claudia).

It can be observed that the average value of cholesterol decreased from 187.79 to 168.25. Before the treatment they were 28% above the recommended value and after it was approximately 20%, although in this situation with values not so close to the allowed value. Therefore it is observed that the applied diet gives positive effects in the decreasing of the cholesterol level (Refugio).

Dieta y Colesterol

	Condicion		Row Summary
	ANTES	DESPUES	
Nivel_de_colesterol	24	24	48
	137	125	125
	163.5	146.5	153.5
	179	165	172.5
	213.5	185.5	199.5
	250	215	250
	187.79167	168.25	178.02083

S1 = count ( )
S2 = min ( )
S3 = Q1 ( )
S4 = median ( )
S5 = Q3 ( )
S6 = max ( )
S7 = mean ( )

Table 3. Cholesterol before and after of the diet

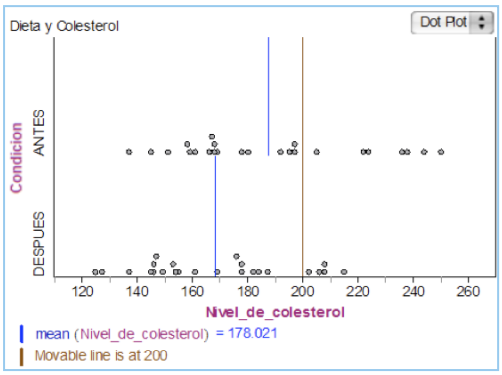

Figure 4. Dot plot with cholesterol levels

In addition to looking at the average (media and median), Claudia focus her attention in the quantity of patients that diminished their cholesterol levels to less than the recommended level, before and after the diet, which is an indicator that they are taking into consideration descriptors in addition to the averages leaning on dot plots. Also observe that both professors use numeric descriptions that are contained in the table and do not use the information in an important graph they have constructed and it allows them to respond to both questions at the same time from the same activity.

CONCLUSIONS

The data analysis and the reasoning developed by the teachers were based principally in the construction of tabular representations with average calculations and in dot plots. We consider that the experience of the teachers with the data analysis focused in descriptive statistics, influenced so that they did not take into consideration important graphic representations like the box diagram, that it was even unknown to some of them, and descriptive measurements such as quartiles, for a more complete analysis of data. The above indicates the is necessary for the teachers to know the potential of a computer environment to go from descriptive statistics in pencil and paper environment to the exploratory data analysis that emphasizes primarily the use of graphic representations and that goes beyond the calculation of descriptive measurements.

REFERENCES

Ben-Zvi, D. (2000). Toward Understanding the Role of Technological Tools in Statistical Learning. *Mathematical Thinking and Learning, 2*(2). Mahwah NJ: Lawrence Erlbaum Associates, Inc., 127-155.

Chance, B., & Rossman, A. (2006). Using simulation to teach and learn statistics. *Proceedings of the 7th International Conference on Teaching Statistics.* ISI-IASE.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). *The Role of Technology in Improving Student Learning of Statistics. Technology Innovations in Statistics Education, 1*(1). Online: repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2.

Cobb, P., & McClain, K. (2004). Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning. In D. Ben-Zvi & J. Garfield (Eds.) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (375-395). Springer.

Curcio, F. R. (1989). *Developing Graph Comprehension*. Reston, VA: NCTM.

Eichler, A. (2008). Statistics Teaching in German Secondary High Schools. In C. Batanero, G. Burrill, C. Reading & A. Rossman, *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.

Garfield, J., & Ben-Zvi, D. (2008). Creating Statistical Reasoning Environments. In J. B. Garfield & D. Ben-Zvi (Eds.), *Developing Students' Statistical Reasoning* (pp. 91-114). Springer Science+Business Media.

Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. Journal *of Statistics Education.* Online: www.amstat.org/publications/jse/v10n3/garfield.html.

Inzunsa, S., & Juárez, J. A. (2007). Evaluación de Cultura y Razonamiento Estadístico: Un estudio con profesores de preparatoria. *Memorias de XII Conferencia Interamericana de Educación Matemática.* Querétaro México.

Konold, C., & Miller C. (2005). *TinkerPlots: Dynamic Data Exploration Software*. Key Curriculum Press

Moore, D. S. (1992). Teaching Statistics as a Respectable Subject. In F. Gordon & S. Gordon (Eds.) *Statistics for the Twenty-First Century*. MAA Notes (26), Washington, DC. Mathematical Association of America, 14-25.

Pea, R. (1987). Cognitive Technologies for Mathematics Education. In A. Schoenfeld (Eds.) *Cognitive Science and Mathematics Education* (pp. 89-122). Lawrence Erlbaum Associates Publishers.