LEARNING BASED ON REAL CONTEXT PROBLEMS AND NOTIONS OF PROBABILITY DISTRIBUTIONS AND EXPECTED VALUE

José Armando Albert Huerta, Juan Ernesto Colunga Cavazos, and Juan Antonio López Esquivel
ITESM, Campus Monterrey, México
albert@itesm.mx

The following investigation explores the learning potential of college students, when using the technique of Problem Based Learning. Particularly it is directed towards probabilistic frequency, probabilistic distributions, and expected value notions. To achieve this, a scenario was designed. It was worked by students, in groups of four or five, in order to investigate relevant information about the given situation. Students were expected to relate the situation with real data. It could be verified that the students based much of their analysis and conclusions on the use of graphic representations. This work also reports on the role of the Exploratory Data Analysis mainly through the graphic representations, in which the students can get a possible solution to a problem of the scenario.

BACKGROUND

The following work is in an emergent area of the Didactic of the Statistics called Exploratory Data Analysis.

On a college level, the teaching of Statistics seems affected by the continuous demand of the professionals. This demand is not just for the use of data on many different conditions and quantities, but for the use of technological tools that make the calculus and interpretations easier. The gathering of data is usually the result of procedures that involve a monetary, time, and people cost. However, most of the time, the traditional statistic perspective does not take advantage of data. From this perspective, data have been used to estimate parameters, test hypothesis, and make predictable models. Nevertheless, on recent years, mainly with John Tukey (1962), it has been considered that data can reveal more useful information.

On recent years the exploratory data analysis has been more accepted. A big part of this growth is because of the availability of computers. They have made possible the development of new graphic methods that support exploratory data analysis as an effective philosophy. This is because the analysis trusts sample data, makes little suppositions of data's structure, emphasizes the description of models more suitable to the data, and is used for working with statistic techniques less complicated. The analysis also allows the user makes decisions not only through models and relations, but through studying the effects behind the relations.

From what is mentioned above, an interest to take exploratory data analysis to the classroom, or at least, to develop cognitive abilities in the students around it, takes place.

On didactic of statistics, there are various efforts towards this direction. Batanero and Godino (2001) announce some educative characteristics of the exploratory data analysis: possibility to generate learning situations referred to topics of interest to students; strong support on the graphic representations for the use of multiple representations of data, without needing a complex mathematic theory. However, the educational research about exploratory data analysis still remains incipient.

CONTEXT

The teaching of statistics has also been affected by the appearance and use of learning techniques, which are a result of the psychology learning advances and pedagogy. One of the most relevant techniques is Problem Based Learning (PBL). This learning technique, which was developed at the teaching of Medicine, now has been extended to all knowledge areas, and has being adopted by different institutions of university level, for example: Maastricht University, Holland; Wheeling University, West Virginia, United States, and Tecnológico de Monterrey, Mexico. Our investigation is developed into the context of PBL technique with engineering students of the last institution mentioned. With this technique, we want students to become responsible of his own learning; assume a participative and collaborative role in the process through certain activities; have contact with his surroundings; compromise himself on a process

by thinking on what their are doing; develop autonomy and use technology as a helpful tool to enrich his apprenticeship.

METHODOLOGY

This study proposes to explore the apprenticeship's potential of college students when they use the PBL technique. Mainly we focus upon the notions of accumulative probability distribution and expected value. The work method was to form groups of four and five students with a specified assigned role to each of them (leader, secretary, tracker, technology expert, and journalist) to work together for three class sessions. The students worked on a scenario with real data. For them, the purpose was to construct a possible solution based on the rubrics, which be established to evaluate their work.

With this technique, students usually follow next steps:

Step 1: *Read and analyze the problem's scenario.* For students the purpose is to verify their scenario's comprehension through discussion within the group.

Step 2: *Make a list of hypothesis or ideas.* Students usually have theories or hypothesis about the causes of the problem; or ideas of how to solve it. These ideas or hypothesis have to be listed. They will be either accepted or rejected while the problem was been solving.

Step 3: *Make a list of what is known.* A list must be made with everything the group knows about the problem or situation.

Step 4: *Make a list of what is unknown.* A list of everything the group thinks they should know in order to solve the problem must be made. Different kinds of questions could be adequate; some of them may be related with concepts or principles that must be studied in order to solve the problem.

Step 5: *Make a list of what has to be made to solve the problem.* Plan the investigation. Make a list of the actions needed to be done.

Step 6: *Define the problem.* The definition of the problem consists on writing down some statements that explain clearly what the group is resolving, producing, responding, probing or demonstrating.

Step 7: *Gather information.* The group localizes, stores, organizes, analyzes, and interprets the information from different sources.

Step 8: *Present solutions.* The group presents a report or makes a presentation that shows the recommendations, predictions, or any other aspect relevant to the solution of the problem.

The scenario consisted on a problem situation referred to the maximum percentage of alcohol that a local government must allow before applying a fine for driving. The scenario is the next one:

> *Scenario: Death or life situation.* The Local Congress of Monterrey, before approving or rejecting the proposal of the governor, wants to know the amount of alcohol (limit value) allowed drivers before prohibiting driving. The Congress recognizes that alcohol is getting a bigger problem in our society; then they are planning to create a commission that study the problem and present a proposal to them.
>
> You and your group are that commission.
>
> The Congress asks you to compare the results of the accidents caused by alcohol in Monterrey with other cities. You also have to obtain the percentile of limit value of alcohol in blood in relation to the number of deaths caused by alcohol in car crashes. You have to give some conclusion and complement your report with important reasons that invite the community to reflect on the consequences of drinking alcohol and driving.

RESULTS

Use of Technology

In general, it was observed the students have a great ability to find relevant data about the problem: by Internet, or in a consultation with the government. Independently of the assignation of the roles, all students look for information. Some of them personally consult government offices. Their search included official internet pages with data about medical, traffic legislation

and socio-economical aspects of potential countries mainly. The students complained of the lack of national information.

Difficulties Presented During the Process

Some students obtained a reasonable solution based on their analysis, but their variable of study was not clearly defined. They managed it, but without establishing the variable explicitly. Some of them preferred to work only on quality variables maybe because there is plenty information on their sources. Others did not get to a graphic description of data, staying on numeric and table's descriptions of the percentage type. That works were the most limited works. An important confusion of the graphic description of quality data was to call histogram to a bar diagram. It seems all graphic with vertical rectangles is a histogram for some students.

The unity of measure of the quantity of alcohol was another difficulty that the students had. Information sources use different unities like grams per liter or milligrams per deciliter, or percentage of alcohol, etc. So it was difficult that the students compared quantities of alcohol. It seems like another obstacle was the lack of data about accidents with alcohol on the blood. They said that they needed these data and if they did not find them, they separated from a better proposal. In general, the assigned time of a week to finish the work was not enough because many of the students asked for another week to expose their work.

Students' Solution Strategies

Although some students tried to leave planning for outside the classroom, exhortation was given to them to do it in class. A guide questions of the technique were an important help. The students tended to base their analysis of graphs already made by other people, but they encouraged themselves to interpret them. The ones that elaborated their own graphs from the data that they found were the ones nearer to an acceptable solution. At the end of their analysis, they thought that the information was applicable to any place.

Description of the route what is based on accumulative and percentile frequency distribution. It was observed that those students who worked with relative frequencies and relative accumulative frequency histograms had an initial deduction of the necessity to find a statistic relation between accidents and the quantity of alcohol in the blood. It looks like making the histogram of accumulative frequencies in percentage is what makes them visualize the most viable solution. Students were not asked to elaborate those graphs, only to make graphic and numeric descriptions. Teacher suggested their students manage percentiles. The students showed to improve their understanding about the meaning of percentiles.

Description of the route what is based on the concepts of frequency probability and probabilistic distribution. Mostly students managed percentages in their works. Those with the most acceptable solution pointed out the necessity of connecting the probability of accidents with the quantity or levels of alcohol. It looks like this took them to a search of effective data, and to the elaboration of valid histograms.

Description of the route what is based on the concept of expected value. An error on the concept of average was detected because the group averaged the frequencies instead of the alcohol concentrations. Other groups, in order to propose limit value, chose the value they obtained when they averaged the limit values of different countries, maybe because of the idea of representative of average.

DISCUSSION

According to this experience, the use of the didactic technique PBL promotes abilities around the use of technology like searching information, presentation of data through different graphs and communication skills.

The study of real data exploration resulted very rewarding for the students: it motivated them, and aroused the creativity on the search of solutions. Also, the problem of alcohol and transit accidents collides on a collateral way the ethic and social questions of driving with certain alcohol level in the body.

The students felt graphics and direct use of data were a great intuitive support to conjecture explanations and possible solutions.

The activity allowed the students to have a better comprehension of different types of data that they found when they approached official statistics of vehicle accidents. Carr (2002) states that data classification is still debated. However, the students tried classifying nominal, ordinal, and interval variables, according to Stevens' proposal (1946).

The notion of expected value was presented as the average of the limits of alcohol on different countries. Graph of accumulated frequency was the most used, but not yet with all the potential that it has.

The activity resulted as a good reference to reaffirm concepts like the type of data, random variables, probabilistic distribution and probability of frequencies.

REFERENCES

Batanero, C. (2001). *Didáctica de la Estadística.* Granada: Grupo de Investigación en Educación Estadística.

Batanero, C. and Godino, J. D. (2001). *Análisis de datos y su didáctica.* Materiales para la asignatura. Departamento de Didáctica de la Matemática. Universidad de Granada.

Carr, R. (2002). A data analysis tools that organizes analysis by variable types. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Stevens, S. S. (1946). On the theory of scales and measurement. *Science*, 103, 667-680.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33, 1-67.