INTERPRETATIVE SKILLS AND CAPACITY TO COMMUNICATE STATISTICALLY: A DIFFERENTIAL ANALYSIS

Pedro M. Campos[1] and Emília Oliveira[2]
[1]Statistics Portugal and LIAAD/FE, University of Porto, Portugal
[2]Escola Secundária de Tomaz Pelayo, Santo Tirso, Portugal
Pedro.campos@ine.pt

One of the common dilemmas in the teaching and learning of statistics is the (lack of) reasoning that students demonstrate to deal with data. It is indeed a major problem in Initial Data Analysis: there is not a clear understanding of the data in order to highlight useful information and suggest conclusions to support decision making. To better understand the types of problems involved in this dilemma we have made a survey, based on simple problems in which students are invited to interpret the original data as well as tables and graphics. The main results of this survey indicate that students have a need of creativity to interpret statistical data, particularly when the concepts are learned in previous levels. The consolidation of the statistical concepts is missing as there is not enough memorization by students in order to apply the descriptive facet of the statistical cognition.

INTRODUCTION

The Curriculum and Evaluation Standards for School Mathematics, published in 1989 (NCTM, 1989), introduced rules on the Probability and Statistics for all levels of education. On this basis, the current guidelines for the teaching of mathematics recommend a strong development of content, concepts and procedures which should acquire a greater degree of expansion over the years of schooling, so that at the end of secondary school students have a sound knowledge of basic statistics. In order to understand the fundamental ideas of statistics, students should work directly with data. Data analysis and the study of probabilities provide a natural environment for students to set up connections between mathematics and other subjects and their everyday experiences. In fact, in recent years, the teaching and learning of statistics has become more practical, and interactive. One of the main causes that contributed to this improvement is that the paradigm of statistics pedagogy has changed in the last two decades: statistics is more than a branch of mathematics supported by data analysis: it involves experience planning and problem-based matters that needs careful thinking and reasoning (Garfield, 2002; Garfield & Gal, 1999; Moore, 1992, 1998). A very good relation between statistical reasoning and statistical cognition is made by Maromet al. (2008) who emphasize that statistical reasoning concerns the mental processes which shape the process and representation of statistical cognition. In that study, the authors define statistical cognition as involving several aspects such as normative, descriptive and prescriptive. In this sense, the descriptive facet of statistical cognition is the process of how people use statistical knowledge.

The descriptive facet of statistical cognition is the process of decision making. Having in mind the amount of information available, this process is very complex nowadays. For example, given a set of data, often students do not know how to use the basic measures of location to determine the center of the sample or what kind of graphics should use in order to better summarize the information. It is indeed a major problem in Initial Data Analysis: there is not a clear understanding of the data in order to highlight useful information and suggest conclusions to support decision making. This is an important challenge to students and to all citizens in general.

Taking into account the importance of the statistical reasoning, this paper concerns the use of statistical knowledge by a set of students from different levels of education. We have made a survey, based on simple problems in which students are invited to interpret the original data as well as tables and graphs. The questionnaire was applied and adapted to different education levels from the 3rd cycle of basic school to higher education. The main results of this survey indicate that students have a lack of creativity to interpret statistical data, particularly when the concepts are learned in previous levels. When brought to the need to write a text to illustrate a data set, the situation is even more dramatic, as responses demonstrate a lack of statistical reasoning. In the case of higher education the gap is even greater, with students having serious difficulties in making a histogram to represent solid figures. A set of initiatives is proposed in order to overcome these

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

problems. This is a part of a greater study where we aim at defining the most important components of one main aspect in the teaching and learning of any subject: the *consolidation of the concepts*.

THE SURVEY

In Portugal, the syllabus of Mathematics contains topics of statistics that are taught since the 1^{st} level of basic school. The program of Statistics, starting in the 1st level of basic school until the secondary school is described in Table 1.

Table 1. Topics of statistics taught in Portuguese curricula in basic and secondary levels

Basic 1^{st} cycle	**Basic 2^{nd} cycle**	**Basic 3^{rd} cycle**	**Secondary**
Levels 1 to 4 *(students aged 6 to 9)*	Levels 5 and 6 *(students aged 10 to 11)*	Levels 7 to 9 *(students aged 12 to 14)*	Levels 10 to 12 *(students aged 15 to 17)*
Graphing and interpretation of data; random situations	***Data interpretation***	***Statistical planning***	***Statistics and Probability***
>Reading and interpretation of data in tables and graphs	>Problem formulation	>Problem specification	>History of Statistics, Census, Surveys, Sample issues
>Data classification using Venn and Carroll diagrams	>Types of data	> Data collection	>Types of data (qualitative, quantitative), Random variables, Cumulative function;
>Frequency tables (absolute)	>Frequency tables (absolute and relative)	>Population and Sample	> Discrete and continuous variable
>Scatter plots, pictograms,	>Bar charts, Pie charts and steam-and-leaf charts	> Data Organization, analysis and interpretation	>(More) Location and dispersion measures
>Bar charts	>Arithmetic mean	>Histograms	>Bivariate distributions; Correlation; Regression
>Mode	>Outliers and range	>Location and dispersion measures	>Random experiments; Probability theory: Axiomatic and theorems; Conditional probability
>Random situations		> Results' discussion	>Probability distributions: Binomial, Normal.
		>Probability	>Combinatory analysis
		> Random event, and random experience	

The curricula contain several topics that aim at giving students more interpretative skills and capacity to communicate statistically. Furthermore, in the university, some courses include Statistics in their curricula. Depending on the type of course, in general, statistical topics range from Descriptive Statistics and Data Analysis to Probability, Random Distributions and Statistical Inference.

The whole sample of the survey includes 647 students ranging from the 7^{th} level of basic school until the 1^{st} grade in the university (see Table 1 for the corresponding ages and school levels). The sample was split into two periods: in 2008, the survey was applied in Period 3 (last week of the school year, in June) while in 2009 the survey was applied in Period 1 (December). This division was made in order to capture the level of knowledge in Statistics acquired by students according to the time where statistical concepts were introduced, as we describe later in this paper. The questions of the survey were based on simple problems in which students have to interpret original data as well as tables and graphs. Some questions are of intuitive answer and do not require any calculation. Others have small calculations involving the creation of simple graphics to support the final answer. Some important statistical aspects are involved such as (1) Likelihood, (2) Data visualization, (3) Measures of location (the concept of mean), (4) Free creation of graphs and (5) Communication with statistical knowledge. For a matter of simplicity, and also because it was the only question applied to all levels of education in this study, the analysis presented in this paper, just describes the question 4. In this question a set of numbers was presented, corresponding to the height (in centimeters) of ten football players: 175, 189, 165, 171, 175, 177, 192, 184, 186, and 169. As the variable is quantitative, in this case the histogram is the best way to represent the data. Histograms are introduced in the 8th level, as pointed in Table 1.

A DIFFERENTIAL ANALYSIS

Question 4 is an open question for which the main results are presented in Table 2. Answers to this question have been classified *a posteriori* according to the following scale: 1) Scatterplot, line graph, no answer; 2) Bar charts; 3) Histogram (right answer).

Table 2. Main results of question 4

Level		Question 4 - Period 3 (2008) 1	2	3	Total (2008)	Question 4 - Period 1 (2009) 1	2	3	Total (2009)
SUP*	Count	21	19	10	50	7	13	5	25
	% within Level	42,00%	38,00%	20,00%	100,00%	28,0%	52,0%	20,0%	100,0%
7	Count	1	18	0	19	28	16	0	44
	% within Level	5,30%	94,70%	0,00%	100,00%	63,6%	36,4%	,0%	100,0%
8	Count	8	9	7	24	12	30	22	64
	% within Level	33,30%	37,50%	29,20%	100,00%	18,8%	46,9%	34,4%	100,0%
9	Count	45	26	2	73	13	24	13	50
	% within Level	61,60%	35,60%	2,70%	100,00%	26,0%	48,0%	26,0%	100,0%
10	Count	14	10	3	27	63	9	1	73
	% within Level	51,90%	37,00%	11,10%	100,00%	86,3%	12,3%	1,4%	100,0%
11	Count	31	13	1	45	12	9	17	38
	% within Level	68,90%	28,90%	2,20%	100,00%	31,6%	23,7%	44,7%	100,0%
12	Count	27	5	1	33	54	26	2	82
	% within Level	81,80%	15,20%	3,00%	100,00%	65,9%	31,7%	2,4%	100,0%
Total	Count	147	100	24	271	189	127	60	376
	% within Level	54,20%	36,90%	8,90%	100,00%	50,3%	33,8%	16,0%	100,0%

* The SUP level corresponds to undergraduate students in the University (at the University of Porto - Faculty of Economics).

The survey was applied twice in different periods of the school year. This means that for students of the 8th, 9th, 11th and 12th levels, statistical concepts were already taught in both years 2008, and 2009. On the other hand, for students of the 10th level and superior (SUP), there are some differences: in the 2008 sample, topics of statistics were already learned by the students of the 10th level (this situation did not happen in 2009). For the students in the university, the opposite situation happens: in 2008 the statistical concepts were not yet reviewed, while in 2009 the statistical concepts were already reviewed by the teacher.

As we can see in Table 2, the best results are obtained in the 8th level in both 2008 and 2009. This is indeed the level where histograms are introduced in the Mathematics curriculum. That is why results seem to be so uneven: students of future levels, such as 11th, 12th and the superior should, in principle, answer correctly to this question, but many wrong answers are found, mainly in 12th and the superior. This is probably a matter of *concept consolidation:* statistical concepts were taught but not internalized/memorized by students. Statistics is usually not assessed by the mean of an official exam (other topics in Mathematics are) and that is maybe a problem of *unsetting* of the main concepts, particularly when the concepts are learned in previous levels. The Pearson Chi-square test confirms the existence of an association ($p<0.001$) between the school level and the quality of the answers.

CONCLUSIONS AND FUTURE WORK

The main results of this survey indicate that students often fail to interpret statistical data, particularly when the concepts are learned in previous levels. The consolidation of the statistical concepts is missing as there is not enough memorization by students in order to apply the descriptive facet of the statistical cognition.

Only one question of a survey aiming at measuring the capacity of free creation of graphs was analyzed. The results of the survey containing all the questions are to be disseminated in the near future.

REFERENCES

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal Statistics Education*, *10*(3).

Garfield, J. & Gal, I. (1999). Teaching and Assessing Statistical Reasoning. In L. Stiff (Ed.), *Developing Mathematical Reasoning in Grades K-12*, *National Council Teachers of Mathematics 1999 Yearbook* (pp. 207-219).

Marom, R. B., Fidler, F., & Cumming, G., (2008), Statistical Cognition: Towards Evidence-Based Practice in Statistics and Statistics Education, *Statistics Education Research Journal*, *7*(2), 20-39.

Moore, D. S. (1992). Teaching Statistics as a Respectable Subject. In F. Gordon & S. Gordon (Ed.), *Statistics for the Twenty-First Century* (pp. 14-25). Washington, DC: The Mathematical Association of American.

Moore, D. S. (1998). Statistics Among the Liberal Arts. *Journal of the American Statistical Association*, *93*, 1253-1259.

NCTM - National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics.*

NCTM - National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*, Portuguese Version: Normas para o currículo e a avaliação em matemática escolar, Lisboa: APM.