COMMENTS ON THE LEARNING IN INTRODUCTORY STATISTICS COURSES

Marcos Nascimento Magalhães
University of Sao Paulo, Brazil
marcos@ime.usp.br

A recurrent problem in teaching is to evaluate what the students will retain from the contents discussed in class. For the students who completed two basic courses on statistics, we applied a questionnaire with a test, 3 months after the end of the course. The test was composed of 50 true-false questions. The results revealed that students could satisfactorily answer the questions directly related with definitions. However, there was no such performance when the questions required additional procedures related.

INTRODUCTION

There is a growing use of statistics in the world and, in the last years, almost all discussions have included numbers and statistics to support arguments. Though sometimes with wrong or inadequate use, tables and graphs appear frequently in newspapers and magazines in an attempt to give credibility to opinions. The use of the quantitative apparatus makes great difference to support arguments in all fields. Thus, the understanding of basic ideas in the field of statistics collaborates to increase citizenship.

In this paper, we discuss the learning process in introductory statistics courses attended by students of the Institute of Mathematics and Statistics of the University of Sao Paulo, Brazil. The curriculum of Bachelor in Computer Science, Mathematics, Applied Mathematics and Statistics have two semesters of statistics in the first year; classes being a mix of students working in to complete the four degrees. A similar content is required of students who are pursuing a degree in Mathematics Education. The courses include topics as descriptive statistics, probability, discrete and continuous random variables, estimation and hypothesis tests. The objectives of these courses would be to obtain *statistics literacy* and *statistics reasoning* as defined in delMas (2002).

In Brazil the discussion on statistics education is not yet a common theme in our universities. The statistics faculties are, in general, linked to mathematics institutes and, consequently, their main researches are on intrinsic topics of statistics or probability. Also, universities criteria for promotion are usually heavily focused on paper publications and there is an increasing pressure in this direction. For these reasons, faculty members have the tendency to give low priority to undergraduate courses. One of the objectives of this research is to encourage faculty members to discuss their teaching practices on these basic statistics courses. As discussed in Gal and Garfield (1997), the assessment of students' learning and understanding of statistics is itself an important issue in statistics education.

During the first week of the first semester of 2005, we applied a questionnaire to the students that completed, in December/2004, two basic courses in statistics. The three-month interval corresponds to the period of summer vacations. This period helped us to identify the knowledge really learned and that, in some sense, could be used in the future. It was important to apply the test in the first week to avoid running the risk that a new course, attended by the students, might review statistic concepts as part of its initial study content. Students who take the Bachelor Degree in Statistics are certainly included in this case, i.e., they would certainly have a review early in the course of basic content which they would need throughout. However, one cannot tell if other students might not be in the same situation.

The questionnaire is composed of two parts. In the first, students are required to inform the career pursued and their relationship with the subject of statistics after the end of the course work. The second part consists of a test with 50 true-false questions. The response time was not measured but it took around 40 minutes. Teachers and students had previously evaluated the questionnaire and suggestions had been incorporated to the final version. Despite its limitations, a true-false test facilitates the participation of the students. The questions did not require complex computations and it was not necessary for them to use a calculator. We sampled 173 students divided as follows: Mathematics Education (58), Bachelors in Computer Science (34), Mathematics (23), Applied Mathematics (32) and Statistics (26). Each year, the number of students admitted in the careers under study is of 330. However, our population of interest was, approximately, 280 students who had been approved in both statistics courses in 2004.

The set of questions included graph analysis, frequency table, measures of position and dispersion, probability, binomial and normal distributions, point and interval estimation and hypothesis tests. The 50 questions were divided as follows: 19 questions on descriptive statistics and measures for random variables, 17 questions on probability and random variables and 14 questions on statistical inference. The classification took into consideration the preponderant part of each question, since some of them included more than one topic.

DESCRIPTIVE DATA ANALYSIS

The results and comments of our descriptive analysis are based on the score of the test. To aid presentation, we set a maximum of 100 marks in the test, assigning 2 marks for each correct answer. In our calculations, we used the software *Minitab* version 13 and *SPSS* version 10.

For each student we computed the score test inside of each content. Figure 1 presents a box-plot with the results. The mean values are also presented below the label of the subjects. As revealed by the median, the score is better in Descriptive and worse in Inference questions. Performance in Inference is quite symmetrical around the median. The Descriptive subject presents the largest range, however 50% of the higher scores are concentrated in a relatively narrow interval. One possible conclusion is that the students understood better this content. Despite the fact that this can be true, it is important to remember that part of the Descriptive topics rely on common sense, which might occur almost regardless of a statistical learning process. For this reason, we think that the scores on Inference questions are important to identify different levels of the statistics reasoning. In general, grades in the Descriptive part of the course are better and, from the results shown here, the same tendency occurred after the course had ended. In fact, this supports a teaching approach that we have used in our classes and that was also incorporated in a book (see Magalhães and Lima, 2005), i.e.; topics of Descriptive Statistics are included in all parts of the basic statistics courses, taking advantage of the intuition to help students to understand other ideas.

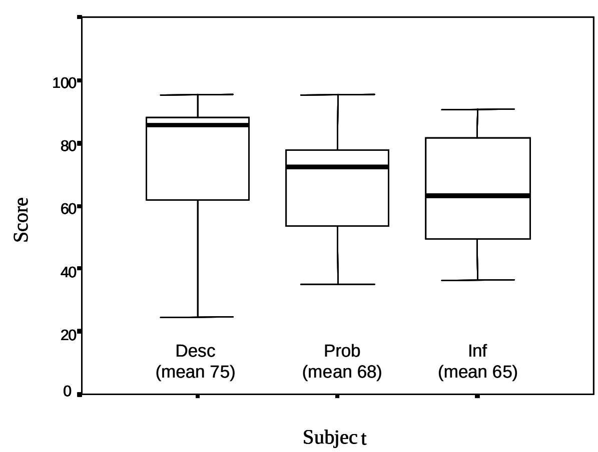

Figure 1: Box plots of the Score by subject

In Table 1, students' performance, through mean and standard deviation, is presented for each career in three different moments. We reported the score obtained in the test, the grade (average) in the two courses attended by the students and the grade in the admission exam. Again, the results were put into a common basis with 100 marks maximum.

According to Table 1, the five careers had almost the same dispersion in the score test. For grades in the course, the dispersion had one relatively high value with students of Mathematics Education. Also, at admission in the university, students of Applied Mathematics were more homogeneous than the others careers. The highest mean score came from the bachelor students of Statistics and Applied Mathematics, even though they were not the students with better past grades (admission or courses). It seems that the motivation can justify this result. These students could be wishing to learn statistics contents more than the other students. The mean grades in the courses were quite similar. However, in the test, the mean score of Mathematics Education students was the lowest value. It was 10 marks below the worst average performance among the bachelor's students. Since they had taken different classes with different instructors, it is possible that the criteria for grading were not the same.

Table 1: Mean (std. deviation) by career

Career	Test	Course	Admission
Math. Education	60 (10)	67 (15)	48 (6)
Computer Bach.	74 (11)	66 (10)	69 (5)
Statistics Bach.	76 (10)	65 (11)	58 (5)
Mathematics Bach.	71 (10)	64 (10)	61 (6)
Applied Math. Bach.	76 (9)	64 (10)	64 (2)

We performed a cluster analysis, see Barroso and Artes (2003), considering the total score obtained in each subject by each student. The objective of the analysis is to identify students with similar performance. The analysis indicated that 4 groups were a convenient choice to divide the students. We presented on Table 2 the mean and standard deviation for the score of each cluster according the subject. Also, we showed on Table 3 the number of students in each career.

Table 2: Mean (std. deviation) by cluster and subject

Subject	Group 1	Group 2	Group 3	Group 4
Descriptive	77 (13)	64 (12)	69 (8)	90 (7)
Probability	77 (8)	54 (9)	57 (10)	84 (9)
Inference	63 (9)	68 (13)	49 (8)	80 (8)

Table 3: Number of students by group and career

Career	Group 1	Group 2	Group 3	Group 4
Math. Education	9	24	24	1
Computer Bach.	16	5	3	10
Statistics Bach.	8	5	2	11
Mathematics Bach.	11	3	5	4
Applied Math. Bach.	10	7	2	13

According to Tables 3 and 4, Group 4 – predominantly composed by students from Computer, Statistics and Applied Mathematics - presented the highest grades. Groups 2 and 3 presented a worse performance. They were mainly composed of students from Mathematics Education. In Group 2, we had low scores in Descriptive and Probability and regular ones on Inference. For Group 3, the low grades were in Probability and Inference and regular for Descriptive Statistics. The intermediate cluster is Group 1, composed by several students from diverse careers. We noted that almost 50% of the students from Computer and Mathematics degrees are in this group.

We also performed a cluster analysis on the questions. The objective of the analysis is to establish set of questions with the same level of difficulty for the students. For each question, we computed the percentage of students that answered the questions correctly and we built the dendrogram with the complete linkage option. From this diagram, we established 4 sets of questions, labeled A, B, C and D. In Table 4, we presented the main descriptive statistics for each set.

Table 4: Descriptive statistics for cluster of questions

	Set A	Set B	Set C	Set D
Frequency	32	9	5	4
Mean	81	56	34	51
Std. Dev.	11	8	5	5
Median	86	58	35	49
Maximum	95	69	38	59
Minimum	56	47	25	48

Set A is characterized by a large number of questions and with high level of correct answers. The Sets B and D show the intermediate scores, while Set C had questions with very low score. Sets C and D have standard deviation equal to 5, the lowest value among the 4 sets. For Set A, the range was approximately 4 times the standard deviation. The median of Set A was higher than the mean, which indicates an asymmetrical behavior around the mean value. Sets B, C and D had values of mean and median quite close to each other.

Looking at the questions from Set C, we realized that they could be considered the most difficult in the test, but they were not really hard questions. They were questions which required the student to extend a little bit the usual concepts presented in class. In a similar way, we can say that the same happened, though with less intensity, for questions from Set D. The questions in Set B needed, in general, a few thinking and/or computational steps. On the other hand, almost all the questions in Set A were direct applications of mathematical expressions and definitions.

FINAL REMARKS

In this paper we evaluated several students that completed the required two basic statistics courses. Our comments are based only on the descriptive analysis of data collected.

The Inference part of the test seemed to offer more problems to the students than the other subjects. Also, as we expected, the most difficulty questions for the students are the ones that require steps further from the basic concept definition. From the set with the 5 most difficult questions, 2 came from Inference and the other 3 required additional thinking.

Since the course syllabuses are similar in all careers, we might explain the worse performance in the test of the Mathematics Education students with a focus on their background. Apparently, the basic statistics courses took by these students were not able to conduct them to a reasonable statistical literacy. This is a challenging situation since these students will be future school teachers.

In general, for all students, we had expected a better performance. They seemed to have difficulty to think one step further from the immediate statistics concepts focused during classes. Maybe this problem is a consequence of the excessive emphasis in technical operations or in manipulation of formulas instead of on the discussion of main ideas of the statistics.

There is a challenge to everyone that, sincerely, tries to teach sound basic statistics courses. The first step toward this end is to open a discussion with the faculty members about these courses and this could be the hardest part of the job.

REFERENCES

Barroso, L. P. and Artes, R. (2003). Análise Multivariada. Presentation a joint meeting 10th SEAGRO and 48th RBRAS, Lavras, MG.

delMas, R. C. (2002). Statistical literacy, reasoning, and learning. *Journal of Statistics Education,* 10(3).

Gal, I. and Garfield, J. B. (Eds.) (1997). *The Assessment Challenge in Statistics Education.* Amsterdam: IOS Press and the International Statistical Institute.

Magalhães, M. N. and Pedroso de Lima, A. C. (2005). *Noções de Probabilidade e Estatística* (6th edition). São Paulo: Edusp.