Poor Results: What is wrong?

Magalhães, Marcos Nascimento
University of Sao Paulo, Statistics Department
Rua do Matão, 1010
Sao Paulo, 05508-090, Brazil
E-mail: marcos@ime.usp.br

1. Introduction

The Institute of Mathematics and Statistics of the University of Sao Paulo in Brazil offers degree in five careers, four are degrees of bachelor: Computer Science, Mathematics, Applied Mathematics and Statistics. The fifth career, Mathematics Education, is offered to students that want be mathematics teachers in middle and high schools. All careers have two basic statistics courses including topics such as descriptive statistics, probability, discrete and continuous random variables, statistics estimation and hypothesis tests. There are slight differences among the Bachelor's courses and the Mathematics Education ones but, generally speaking, their objectives would be to obtain *statistics literacy* and *statistics reasoning* as indicated in delMas (2002). In this paper, we discuss students' main conceptual difficulties in these courses. What it is presented here is part of a larger study that, since 2005, evaluates students' learning in basic statistics courses.

Data for the study is collected, approximately three months after course ended, with a test true-false organized in 50 questions. Despite its limitations, a true-false test facilitates the students' participation. The test is organized with are conceptual and non-conceptual questions. They cover the topics of the two courses and it did not require complex computations neither the use of a calculator. 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.

Table 1 presents a summary of the test application in the last three years. The last row shows the overall results. The average of correct answers for each career is given in a 100 points scale and the respective standard deviation is between parentheses.

Table 1: Summary by career.

Career	Admission number	Test 2005		Test 2006		Test 2007	
		Freq.	Mean (Std)	Freq.	Mean (Std)	Freq.	Mean (Std)
Math. Education	150	58	60 (9.8)	45	67 (11.3)	54	75 (8.9)
Computer Bach.	50	34	74 (10.2)	25	72 (9.0)	31	78 (7.5)
Statistics Bach.	30	26	76 (9.7)	17	74 (11.1)	12	74 (10.5)
Mathematics B.	30	23	71 (10.2)	9	62 (13)	6	79 (11.4)
Applied Math. B.	70	32	76 (9.0)	31	68 (12.2)	19	74 (10.0)
All careers	230	173	70 (11.8)	127	69 (11.5)	122	76 (9.0)

Besides the students' participation have decreased from year to year, we still believe that the test gives a good picture of the situation. It is worth to mention that the participation in the test is voluntary but it is required that the student had succeed in both statistics basic courses. In 2007, the overall average was 76, the best performance ever, however it can not be considered a good score if we realize that the random answer would produce 50 points. In this paper, our aim is to search possible reasons for the poor results obtained in these tests. The discussion relating careers, subjects and backgrounds is present in Magalhães (2007). For a reference on assessing student's learning in statistics consult Gal, I. & Garfield, J. B. (1997).

2. Discussion

In the three years application of the test, the programs of the courses did not change, but the requirements to be approved and topics' emphasis could vary from one class to another depending on the teacher appointed. Despite these differences, it is possible to use the results of all the students together, to establish considerations on the learning situation.

In this paper, we looked at the 5 questions with worst scores. This gave to us an idea of the students' difficulties. One interesting coincidence was to obtain the same set of questions with less percentage of right answers in all years of test application. In Figure 1, the questions focused on are presented, with their original numbers. We also include a column with the correct answers and its respective average percentage value in the period 2005-2007.

7.	Suppose the mean salary in São Paulo is bigger than the mean salary in Rio de Janeiro. Choosing, at random, one worker in each city, is it more probable the worker from São Paulo to receive a better salary than the one from Rio de Janeiro?	Yes ( ) No (X) 39%
11.	From a sample we obtain [0.14; 0.16] as a 95% confidence interval for the proportion of votes of a candidate. Then, with probability 0.95, will this candidate receive between 14% and 16% of the votes?	Yes ( ) No (X) 36%
18.	We present a histogram with a sample of the heights of students. The interval from 160 to 170 cm has the highest percentage. We ask the following: Must the most frequent height of the students in the sample be between 160 and 170 cm?	Yes ( ) No (X) 24%
27.	We present a frequency table from a sample of ages (in completed years). The interval from 36 to 40 indicates 40% and it is the highest percentage in all table. Then we ask: Is the highest frequency age, certainly, between 36 and 40?	Yes ( ) No (X) 34%
46.	A machine is set to produce pieces with 2 cm of diameter. However variations can happen and we will assume that the diameter follows a Normal distribution with mean 2 cm and variance 0.09 cm^2. To verify if the machine is set correctly, a sample of 100 pieces were collected obtaining 2.1 as a sample mean. If μ is the population mean, then is the convenient alternative hypothesis $H_a: \mu > 2$?	Yes ( ) No (X) 40%

Figure 1: Questions with low percentage of right answer.

An initial curious observation is that in all these questions the correct answer was "No". In the test, there was equal number of correct answers with "Yes" and "No", that is 25 questions to each option. We do not have any thought why this happened and we will not focus on this here.

Questions 18 and 27 are related to the concept of representation of observed values by frequency tables and histograms. The students apparently mixed up mode with modal interval. For continuous variables, this point is a bit more slippery than for discrete variables. Identical values can also happen in continuous variables since our observation is always in a specific decimal precision. A point that could have confused the students is the concept of mode of a continuous model variable with one of mode of continuous observed variable. For sure this is a delicate point since that is common to use observed values to create models. So the poor results in Question 18 could have been inflate by these details. Question 27 is related to a discrete observed variable and it is not difficult. Its low percentage highlights the conceptual misunderstanding of representation by frequency tables.

Question 11 deals with confidence intervals. The low scores in this question shows the difficulty of the students to interpret correctly the interval estimation. In Question 9 we presented the correct interpretation in a theoretical and abstract language and we ask if it was corrected. The average right answer was 76%, higher than two times the percentage right answers in Question 11. This reinforces that students could recognize a correct definition but were unable to apply it in a practical problem. This is a challenging situation since the aim of the classes is to make possible the relationship theory-practice.

Variability is the subject of Question 7. The issue that the mean does not tell everything, it was probably mentioned by the teachers during classes. However, again, the application of the concept brings difficulties for the students.

Question 46 checks if the students can formulate, correctly, alternative hypothesis in a statistical test. The important issue here is the comprehension that there is relative independence between the formulation of the hypothesis and the sample results that will be used to decide about the hypothesis of the statistical test. We say relative independence because no one will formulate the hypothesis without a minimal knowledge of the behavior of the parameter. The previous knowledge set the hypothesis and the sample information guides the decision on what hypothesis has statistical significance.

Looking at the effect of the career in the students' performance, we decided to concentrate our discussion in two of them, Mathematics Education and Computer Bachelor, since these are the careers that have, respectively, the minimum and the maximum average score in the admission exam to the university. This exam includes Mathematics and other knowledge areas students had faced on in high school and it can be used as a measure of a background of the students that come to the university. Table 2 presents the average score of the admission exam for each year and career. We also repeated the average score in our test for further comments. The scores are in a same scale of 100 points. Clearly, the Mathematics Education students began university with lower background than the Computer Bachelor ones. From year to year, there is not a big difference among the average admission scores in each career, however, for the average test score, the students of Mathematics Education had an important increase from 2005 to 2007. The Computer Bachelor's students keep relatively high scores in the test trough out the years .

Table 2: Averages for Admission exam and Test score.

Career	2005		2006		2007	
	Admission	Test	Admission	Test	Admission	Test
Math. Education	48	60	50	67	49	75
Computer Bach.	67	74	71	72	65	78

Considering the 5 questions with the worst performance mentioned before, that is, Questions 7, 11, 18, 27 and 46, we compared the two careers. In Table 3, we mark with a bullet if the question is in the lower score group of the respective career and year. Question 18 is the only that appears in all careers and years. Question 11 is not one of the five worst questions for Mathematics Education in any year. Questions 7, 11, 18 and 46 are, since 2005, questions with low scores for the Computer students group. This indicates that the correspondent statistics courses are not addressing properly the concepts relative to these questions. Even though this group has better scores in general, they apparently fail to understand the concepts of the relationship between mean value and distribution, the representation of observed values through histograms and the construction of hypothesis tests.

Table 3: Presence in the group of worst questions by career and year.

Question	Mathematics Education			Computer Bachelor		
	2005	2006	2007	2005	2006	2007
7	•		•	•	•	•
11				•	•	•
18	•	•	•	•	•	•
27	•	•	•			•
46	•	•		•	•	•

3. Final Comments

As an instrument to assess the student's learning in statistics basic courses a test with 50 true-false questions has been applied since 2005. The results indicate that some basic concepts seem to be misunderstood by the students of all careers of the Institute of Mathematics and Statistics of University of Sao Paulo. The poor performance, revealed by the relatively low average scores, pointed out to the need of a search for what may be wrong. In our opinion, it is necessary to strength the connections between theory and real problems. Without this, the concepts for the students will be closed in a box waiting for a key. It is important that our students could think independently and apply statistical concepts without the need of someone else to tell them what to do.

REFERENCES

delMas, R.C. (2002). Statistical Literacy, Reasoning, and Learning. *Journal of Statistics Education* Volume 10, 3.

Gal, I. & Garfield, J. B. (1997). *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press.

Magalhães, M. N. (2007). We think they have learned. Are we right? IASE Satellite Conference on Assessing Student Learning in Statistics, Guimarães, Portugal.

ABSTRACT

The objective of this paper is to discuss students´ learning in basic statistics courses attended by undergraduate students of Mathematics and Statistics Institute of University of Sao Paulo (IME-USP), São Paulo, Brazil. The discussion is based on the results of a test applied three months after the courses have ended.

The data collected included students of the following careers: Bachelor in Computer Science, Mathematics, Applied Mathematics and Statistics and Mathematics Education. All the students had had two semesters courses focused on topics such as: descriptive statistics, probability, discrete and continuous random variables, estimation and hypothesis tests. In the last three years, 2005 to 2007, the students approved in the basic statistics courses at IME-USP are asked to answer a test composed of 50 true-false questions. The students´ participation in the test was voluntary.

From the analysis of the 5 questions with worst scores, we identified three contents that students presented difficulty to deal with: the relationship between mean value and distribution, the representation of observed values through histograms and the construction of hypothesis tests. In order to identify differences among the careers, we also looked at the scores in each career. The effect of background of these students was also considered, observing the grades at admission to the university and the previous performance in the statistics courses.