ANALYSIS OF A BASIC STATISTIC COURSE USING ITEM RESPONSE THEORY

Dalton F. Andrade[1] and Marcos N. Magalhães[2]
[1]Federal University of Santa Catarina, Brazil
[2]University of São Paulo, Brazil
dandrade@inf.ufsc.br

The objective of this study is to evaluate students' learning from the contents presented in class. Data from undergraduate students were collected in three consecutive years since 2005 and from five different careers in the area of Mathematics. For a sample of these students, who successful completed the two basic classes in statistics, a test was applied 3 months after the end of the school year. It was composed of 50 true-false items divided as follows: 19 items on descriptive statistics and measures for random variables, 17 items on probability and random variables and 14 items on statistical inference. A two-parameter item response model was considered and 36 out of the 50 items were selected to build a scale measuring students' basic statistics concepts knowledge. Also, we discuss the difference among the students from the five different careers and the suggestions to improve the teaching-learning process.

INTRODUCTION

The students of all careers in the Institute of Mathematics and Statistics at the University of Sao Paulo in Brazil have two semesters of basic statistics. The Institute offers four degrees of bachelors- Computer Science, Mathematics, Applied Mathematics and Statistics- and the career of Mathematics Education, which is offered to students that want be mathematics teachers in middle and high schools. The courses of Statistics include topics as descriptive statistics, probability, discrete and continuous random variables, estimation and hypothesis testing. 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).

Data for this study was collected in 2005, 2006 and 2007, approximately three months after the second course ended, with a true-false test organized in 50 items. We had a total of 422 participants of a universe of around 690 students enrolled in these three years. Despite its limitations, a true-false test facilitates the students' participation. The test was organized with conceptual and non-conceptual items. They cover the topics of the courses and it did not require complex computations, neither the use of a calculator. The 50 items were divided as follows: 19 items on descriptive statistics and measures for random variables, 17 items on probability and random variables and 14 items on statistical inference. The classification took into consideration the preponderant part of each item, since some of them included more than one topic.

The present paper uses item response theory to create a scale for the basic statistics knowledge obtained after the two courses of statistics mentioned before. Also we compare the level reached by students of each career. The work reported in Magalhães (2007) used the same data set to analyze the items with worst scores and to compare careers.

METHOD

We define the binary random variable U_{ij} with values 1 or 0 according to, respectively, the individual j answers correctly, or not, the item i of the test. Let θ_j be the level of basic statistics knowledge of individual j and we consider the two parameter logistic item response function to characterize the probabilistic structure of U_{ij}. The two parameters are the "discrimination" and "difficulty" of item i represented, respectively, by a_i and b_i (Lord, 1980; Andrade et al., 2000). A high value for a_i suggests item i will discriminate well between the different students' knowledge and a high value for b_i suggests item i is difficult, which in the context of this study indicates that the particular item will be answered correctly only by students with high level understanding. We consider the following model for U_{ij}:

$$P\,(U_{ij} = 1 \mid \theta_j) = \frac{1}{1 + exp\,[-a_i(\theta_j - b_i)]}.$$

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

Estimation procedure uses marginal maximum likelihood method. In order to obtain better adjust, an item *i* was initially discarded if a_i < 0.4 and, in the sequence, for a_i < 0.5. As a final exclusion criteria, we use b_i < -4 and *ep(b)* >1 (standard error greater than 1) to discard item *i*. This procedure results in the discard of 14 items. Using the estimates of the item parameters as the true values of them on the scale with mean 0 and standard deviation 1, empirical Bayesian estimates are found for the basic statistics knowledge for each respondent on this scale (Baker and Kim, 2004). These estimates are rescaled to have mean 50 and standard deviation 10.

RESULTS

The calculations described in previous section were performed using the software BILOG-MG (Zimowski et al., 1996) and the results for the 36 items remained, ordered by difficulty, are shown in Table 1, including the items' subject among Descriptive (Desc), Probability (Prob) or Inference (Inf).

Table 1. Characteristics of the Items

Item	Topic	a	B	20	30	40	50	60	70	80
I17	Desc	0.614	-3.635	0.60	0.73	0.83	0.90	0.95	0.97	0.98
I10	Inf	0.655	-3.402	0.57	0.71	0.83	0.90	0.95	0.97	0.99
I36	Prob	0.691	-3.151	0.53	0.69	0.82	0.90	0.95	0.97	0.99
I16	Inf	0.953	-2.789	0.45	0.68	0.85	0.93	0.97	0.99	1.00
I28	Desc	0.912	-2.664	0.42	0.65	0.82	0.92	0.97	0.99	0.99
I25	Desc	0.864	-2.576	0.41	0.62	0.80	0.90	0.96	0.98	0.99
I30	Prob	1.070	-2.477	0.36	0.62	0.83	0.93	0.98	0.99	1.00
I22	Desc	1.273	-2.471	0.34	0.65	0.87	0.96	0.99	1.00	1.00
I12	Desc	1.678	-2.459	0.29	0.68	0.92	0.98	1.00	1.00	1.00
I04	Inf	1.232	-2.351	0.31	0.61	0.84	0.95	0.98	1.00	1.00
I40	Desc	0.884	-2.331	0.36	0.57	0.76	0.89	0.95	0.98	0.99
I21	Desc	1.554	-2.297	0.25	0.61	0.88	0.97	0.99	1.00	1.00
I09	Inf	0.523	-2.292	0.41	0.54	0.66	0.77	0.85	0.90	0.94
I24	Desc	1.150	-2.280	0.30	0.58	0.81	0.93	0.98	0.99	1.00
I31	Prob	0.551	-2.240	0.40	0.53	0.66	0.77	0.86	0.91	0.95
I29	Desc	1.129	-2.212	0.29	0.56	0.80	0.92	0.97	0.99	1.00
I05	Prob	0.783	-2.178	0.34	0.53	0.72	0.85	0.92	0.96	0.98
I23	Desc	1.427	-2.097	0.22	0.53	0.83	0.95	0.99	1.00	1.00
I20	Desc	0.688	-1.984	0.33	0.50	0.66	0.80	0.89	0.94	0.97
I44	Prob	1.135	-1.855	0.21	0.46	0.73	0.89	0.96	0.99	1.00
I49	Inf	1.075	-1.831	0.22	0.45	0.71	0.88	0.95	0.98	0.99
I38	Prob	1.240	-1.342	0.11	0.31	0.60	0.84	0.95	0.98	1.00
I43	Prob	1.271	-1.322	0.11	0.30	0.60	0.84	0.95	0.99	1.00
I37	Prob	1.010	-1.278	0.15	0.33	0.57	0.78	0.91	0.96	0.99
I41	Desc	0.748	-1.215	0.21	0.36	0.54	0.71	0.84	0.92	0.96
I06	Prob	1.182	-1.054	0.09	0.25	0.52	0.78	0.92	0.97	0.99
I08	Desc	0.984	-1.050	0.13	0.28	0.51	0.74	0.88	0.95	0.98
I14	Inf	0.845	-0.992	0.15	0.30	0.50	0.70	0.84	0.93	0.97
I35	Prob	0.769	-0.957	0.17	0.31	0.49	0.68	0.82	0.91	0.95
I19	Desc	1.083	-0.732	0.08	0.20	0.43	0.69	0.87	0.95	0.98
I48	Inf	0.838	-0.580	0.12	0.23	0.41	0.62	0.79	0.90	0.95
I34	Prob	0.837	-0.473	0.11	0.22	0.39	0.60	0.77	0.89	0.95
I45	Prob	0.547	-0.154	0.17	0.27	0.39	0.52	0.65	0.76	0.85
I15	Prob	0.658	-0.072	0.13	0.22	0.35	0.51	0.67	0.80	0.88
I27	Desc	1.431	0.607	0.01	0.02	0.09	0.30	0.64	0.88	0.97
I18	Desc	1.473	1.069	0.00	0.01	0.05	0.17	0.47	0.80	0.95

We chose integers multiple of 10 (from 20 to 80) as levels to classify basic statistics knowledge and we used the probability of correct answer of the item to interpret these different levels. The discussion here applied anchoring levels as defined in Beaton and Allen (1992) and it was also used in Harraway and Andrade (2006). Following Table 1, we identified for each level, an item, or set of items, to characterize the level such as the probability of correct answer has a reasonable jump when we change to previous and next levels. In Table 2, we summarize our findings and we indicate the concepts relate to the items, using "(+)" to indicate that the level requires all the concepts of the previous levels. Note that we omitted the level 80 since it was not reached in our data set.

Table 2. Anchoring items

Scale(50,10)	Items	Basic statistics concepts
20	I17	Common sense
30	I12, I04	(+) Mean and variance, sample mean
40	I38, I43	(+) Independence, probability, Normal model
50	I19, I48, I34	(+) Histogram/interpretation, Hypothesis test, Conditional probability
60	I27	(+) Discrete variable/interval representation
70	I18	(+) Continuous variable/representation/ mode

We can interpret the levels of basic statistics concepts as a cumulative sequence of attributes of the students that attended the statistics course. In its initial level of knowledge, 20, the student uses the common sense translating, to statistics items, elementary rules of the day life. The intermediate levels, 30 to 50, require the use of definitions and direct application of formulae. The advanced levels, 60 and 70, require students to be able to interpret graphical representations as a summary and to understand that a specific information maybe available only through the original data set.

The performance according to the careers of the students is presented in Figure 1. The number of students' participants is indicated between parentheses. Applied Mathematics, Computer Science and Statistics bachelors seem to have the better results with respect to the knowledge of basic statistics, while the students from Mathematics Education had the worse. There is only one student in level 20 who is in Mathematics Education career. On the other hand, the largest percent of students in level 70 came from Applied Mathematics.

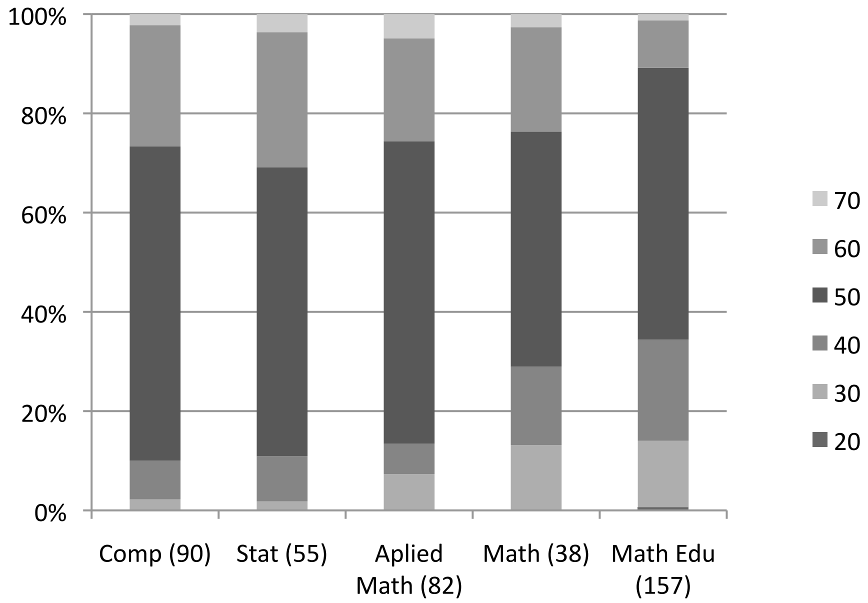

Figure 1. Performance by careers

CONCLUSION

We used item response theory to obtain a scale of basic statistics knowledge. We identified items that can be used as anchoring to different levels. In general, students consider descriptive statistics an easy statistics topic and, curiously, the majority of anchoring items came from this topic. In fact, except for level 40, items on descriptive statistics were present in all anchoring levels. In this way, this fact supports our opinion that descriptive statistics can be a powerful instrument to conduct the learning process toward a complete understanding of basic statistics concepts.

We also studied the performance of the students by career. After the two statistics courses, students from the bachelors of Computer Science, Applied Mathematics and Statistics seem to have reached better understanding than the students from the other careers- bachelor in Mathematics and Mathematics Education. Since the statistics courses are taken in the first year of the university, we suspect that the students' background had important influence in their performance. In Brazil, we have admission exams to the university and some careers attract knowledgeable students than others. Unfortunately, in the last years, teachers' careers were not among them. Consequently, the lower performance of students of Mathematics Education, observed in our data set, can be partially explained by students' insufficient background. Despite this, it remains a challenge for a good education in statistics since the career of Mathematics Education is responsible to prepare the future middle and high school mathematics teachers.

REFERENCES

Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). Teoria da Resposta ao Item: conceitos e aplicações. São Paulo: *Associação Brasileira de Estatistica*. (in Portuguese) *www.inf.ufsc.br/~dandrade/tri*

Baker, F. B., & Kim, S.-H. (2004). Item Response Theory: Parameter Estimation Techniques. Statistics, *Dekker Series of Textbooks and Monographs*. New York.

Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, *17*, 191-204.

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

Harraway, J. A., & Andrade, D. F. (2006). An item response analysis of statistics use in the workplace. *7th International Conference on Teaching Statistics- ICOTS 7.* Salvador, Brazil.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Lawrence Erlbaum Associates Inc.

Magalhães, M. N. (2007). Poor results: what is wrong? *56th Session of the International Statistical Institute.* [CDROM]. Lisbon, Portugal.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). BILOG-MG: *Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Chicago: Scientific Software Inc.