PRE-SERVICE TEACHERS' UNDERSTANDING OF PROBABILITY DISTRIBUTIONS: A MULTILEVEL STATISTICAL ANALYSIS

Theodore Chadjipadelis[1] and Sofia Anastasiadou[2]
[1]Aristotle University Thessaloniki, Greece
[2]University of Western Macedonia, Greece
chadji@polsci.auth.gr

In this paper the problem of pre-service teachers' approaches in solving tasks in probability distributions is discussed. Statistical data were obtained from 98 Greek pre-service teachers from the Western Macedonia University that formed the experimental group and 132 Greek pre-service teachers from Macedonia University that form the control group, enrolled in two basic statistics courses. The control group participates in teacher-centred teaching environment. The experimental group participates in student-centred teaching environment connected with probability distributions. Results indicate the difficulty of control group in defining the difference between discrete and continuous distributions. In addition the estimation of the Poisson distribution parameter λ troubled the majority of the control group while the experimental group showed higher success rate.

INTRODUCTION

Over the last decade, the use of real world projects in introductory statistics courses has increased in popularity. The use of (individual or group) projects for teaching Statistics can, under suitable conditions, help in correcting erroneous perceptions and misapprehensions (Chadjipadelis and Gastaris, 1995). According to Zeleke et al. (2006) real world projects provide students with an opportunity to learn the entire process of a statistical investigation. According to Ponte (1990), working with projects has the biggest advantage of letting the student free to decide what he or she wants to investigate, and presenting the subject matter as a tool for developing the project. The use of projects in introductory statistics has been reported in many literatures (César and Oliveira, 2005). For instance, Chadjipadelis (2003a, 2003b) argued that assigning projects to students sets a frame which encourages interactivity between the instructor and the students, and motivates students to explore the field. Guimaraes et al. (2005) mentioned that projects provide students with a coherent didactical contract, allowing the creation of a secure class climate, to develop autonomy and responsibility as well as their critical sense. Thus this study posed a fundamental question: Does a project improve Greek pre-service teachers' understanding of probability distributions?

PARTICIPANTS-TASKS-DESIGN OF THE PROJECT

The sample consisted of 98 Greek pre-service teachers from the Western Macedonia University and 132 from Macedonia University assumed as experimental group and control group respectively, enrolled in two basic statistics courses. Both groups follow the second year of their studies during the academic year 2008-09. The control group participates in a teacher-centred teaching environment. The experimental group participates in a student-centred teaching environment that includes lectures and projects related with tasks concerning the creation of probability distributions. For the data analysis application a combination of principal components analysis and implicative statistical analysis was used. More specifically, both control and experimental group were taught precisely the same disciplines that connected with basic concepts and probability rules, conditional probability, Bayes law, and finally with discrete and continuous distributions by the same professor. The lectures, the presentation and the formation of the two basic statistics courses, and the demonstrated examples during the courses were exactly the same. Experimental group students during the project had to create real world probability distributions examples/tasks. Those tasks had to be related to their discipline and their future working environment. The tasks were presented afterwards to the following week course by the students. In the final examinations, that took place in June 2009, students of the two groups were examined in the following 6 same tasks. All the tasks were equivalent and the students had in their disposal 3 hours in order to complete their examination. The test consists of following six exercises given below related to the Binomial, Poisson and Normal distribution that required long- answers responses.

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 [© 2010 ISI/IASE]

V1nn: Students rent apartments during their studies. The rents follow normal distribution N(350, 400^2). i)Which is the percentage of the apartments that they are rented 400 euro at least; ii) Which is the percentage of the apartments that they are rented 390 euro tops. V2bb: According to the records of a medicines company, a specific medicine treats 30% of the the cases. If a doctor subscribes this medicine to 4 sick persons find the probability that i) It will be effectual to at least 3 of them; ii) It will not be effectual to anyone. V3pp: The mean number of calls in the secretary desk of a university department during an hour is 12. Estimate the following probabilities: i) There are no calls during 6 minutes; ii) There are more than 3 calls during 30 minutes. Task 4: V4pp. A professor contacts with his students through emails. The emails arrive in his box with a tempo of one message every 6 hours and he corresponds with one message every 8 hours. According to this information, i) Estimate the probability that the professor takes at least 3 messages during a day; ii) If the professor' box is empty estimate the probability to receive 5 messages during a day and he will not the same day. V5bb: A research showed that 60% of the children are overweight. We choose 6 children randomly. Estimate the following probabilities: i) No more than one child is overweight; ii) At least one child is overweight. V6nb: Students' weight (kg) follows normal distribution N(68,100). i) Find the percentage of students that they weigh above 75 kg; ii) We choose 5 students randomly, find the probability that only one is weight above 75 kg.

RESULTS

Results for the experimental group

The results showed differences in the ability to handle every type of given distribution. Table 1a shows the success rates of the experimental group in all types of tasks. It is clear that the task V1nn is favoured related to normal distribution, as it presents the greatest success rate. While the application of the type did not confuse students tasks V4pp and V3pp 3 (4.1% and 12.2% respectively) seemed to be the most difficult related to the estimation of Poisson distribution parameter λ of the asked probabilities.

Table 1a. Total success rates of experimental in all tasks

Tasks	Incorrect	Half correct	Correct
Task 1 (V1nn)	20 (20.4%)	25 (25.5%)	53 (54.1%)
Task 2 (V2bb)	24 (24.5%)	23 (23.5%)	51 (52%)
Task 3 (V3pp)	57 (58.2%)	29 (29.6%)	12 (12.2%)
Task 4 (V4pp)	69 (70.4%)	25 (25.5%)	4 (4.1%)
Task 5 (V5bb)	33 (33.7%)	51 (52%)	14 (14.3%)
Task 6 (V6nb)	26 (26.5%)	51 (52%)	21 (21.4%)

Table 1b. Total success rates of control group in all tasks

Tasks	Incorrect	Half correct	Correct
Task 1 (V1nn)	86 (65.2%)	26 (19.7%)	20 (15.2%)
Task 2 (V2bb)	104 (78.8%)	18 (13.6%)	10 (7.6%)
Task 3 (V3pp)	109 (82.6%)	18 (13.6%)	5 (3.8%)
Task 4 (V4pp)	119 (90.2%)	7 (5.3%)	6 (4.5%)
Task 5 (V5bb)	102 (77.3%)	13 (9.8%)	17 (12.9%)
Task 6 (V6nb)	105 (79.5%)	13 (9.8%)	14 (10.6%)

The KMO=0.738>0.60, measure of sampling adequacy and Bartlett's test of sphericity (sig<0.01), showed the utility of principal components analysis. Table 2a shows 2 uncorrelated factors that explain 74.391% of the total data inactivity.

More specifically variables V1nn, V2bb, V6nb and V5bb are loaded on the factor F1, which explains, following Varimax rotation, 43.559% of the total dispersion. Factor F1 represents the tasks-variables related to Normal and Binomial distribution. This factor highlights the student way of handling Normal and Binomial distribution in a discrete way. Tasks V4pp and V3pp are loaded on the second factor (F2), which explains 30.832% of the total dispersion consisting of the tasks that related to Poisson distribution.

Figure 1a illustrates the similarity diagram of the test. Experimental group responses to the tasks (V1nn, V6nb, V2bb, V5bb, V3pp and V4pp), which are responsible for the formation of two similarity clusters (i.e., groups of variables).

Cluster A consisted of the tasks V1nn, V6nb, V2bb, V5bb representing students' efficiency in solving the problem tasks by using both Normal and Binomial distribution types. The strongest similarity, 0.958602, occurs between variables V1nn and V6nb in Cluster A. It is suggested that pre-service teachers employed similar processes to construct a problem solving strategy estimating a simple Normal distribution probability and a combination of Normal and Binomial probability. The similarity of the tasks V2bb and V5bb is very strong (0.920188) too. The similarity connection of those tasks reveals pre-service teachers' consistency with regard to their performance in evaluating the Normal and Binomial distribution probabilities. On the whole, Cluster A suggests a close connection between the answers (successful or not) to the Normal and Binomial distribution.

Cluster B consisted of tasks V3pp and V4pp and suggests that pre-service teachers employed similar processes to estimate the Poisson distributions probabilities. Cluster A is disconnected from Cluster B, demonstrating pre-service teachers' compartmentalized ways of Normal and Binomial distribution probabilities from Poisson ones. This conclusion was the result of the application both of the analysis to principal components and through implicative statistical analysis.

Table 2a. PCA' Results of experimental group

Experimental group	Factors		
Tasks-Variables	F1	F2	Communality
V1nn	0.881		0.824
V2bb	0.818		0.676
V6nb	0.790		0.655
V5bb	0.680		0.580
V4pp		0.914	0.868
V3pp		0.900	0.860
Eigenvalue	3.302	1.116	
Variance Explained %	43.559%	30.832%	
Total Variance Explained %	74.391%		
Mean score per Factor	0.546	0.219	
Standard Deviation per Factor	0.391	0.321	

KMO= 0.738
Bartlett's Test of Sphericity: x^2=267.737, p=0.000

Table 2b. PCA' Results of control group

Control group	Factors		
Tasks-Variables	F1	F2	Communality
V1nn	0.796		0.643
V2bb	0.793		0.647
V3pp	0.669		0.471
V6nb		0.746	0.561
V4pp		0.741	0.566
V5bb		0.700	0.493
Eigenvalue	1.875	1.507	
Variance Explained %	28.911%	27.446%	
Total Variance Explained %	56.358%		
Mean score per Factor	0.17	0.14	
Standard Deviation per Factor	0.314	0.310	

KMO = 0.621
Bartlett's Test of Sphericity: x^2=91.523, p=0.000

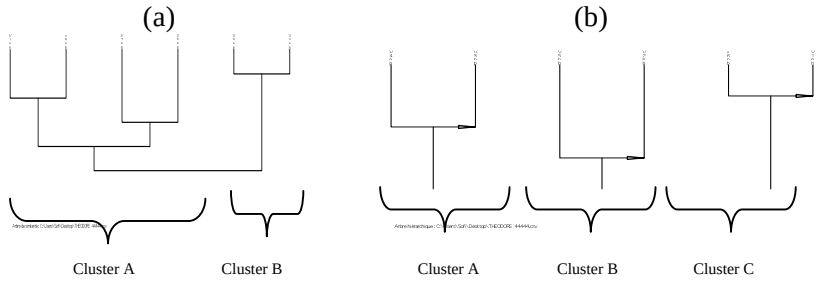

Figure 1. (a) Similarity diagram and (b) hierarchical tree of experimental group

In figure 1b the hierarchical tree shows significant implicative relations between variables of the study. Thus three groups of implicative relationships are identified. The first group of implicative relations refers to variables concerning Poisson distribution (Cluster A); the second group refers to Binomial distribution (Cluster B) while the third one refers to a Normal distribution (Cluster C). This result agrees with the findings that emerged from the similarity diagram.

Results for the Control group: Table 1b shows the success rates of the control group in all types of distributions. Control group success varies across the different kind of tasks. Considering the lowest and the highest percentage, this variation decreases with the kind of distribution: Normal distribution, 10.6-15.2%; Binomial distribution, 7.6-12.9%; Poisson distribution 3.8-5.5%. According to Table 1b, tasks connected with Poisson distribution seemed to be the most difficult as they present the lowest percentages, task V3pp 3.8% and task V4pp 4.5%. Also tasks connected with Binomial distributions, appeared to be difficult as it presents low percentages of 7.6% for task V2bb and 12.9% for task V5bb. Task V6nb which has a combination of a Normal and Binomial distribution seems that troubles the control group (10.6%). Finally, the task V1nn although it has the higher success rate of 15.2%, still this percentage is low.

The KMO=0.621>0.60, measure and Bartlett's test of sphericity (sig<0.01), showed the utility of principal components analysis (Table 2b). Table 2b shows 2 uncorrelated factors occurred, that explain 56.358% of the total data dispersion. More specifically tasks V1nn, V2bb and V3pp are loaded mainly on the factor F1, which explains, via Varimax rotation, 28.911% of the total dispersion. Finally, last on the significance scale for this factor lies the task V3pp. Tasks V6nb, V4pp and V5bb are loaded on the second factor (F2), which explains 27.446% of the total dispersion.

The similarity diagram in Figure 2a shows how tasks are grouped according to the similarity of students' solutions. This fact allows the arrangement of Control group's answers (V1nn, V2bb, V3pp, V4pp, V5bb and V6nb) to tasks into groups according to their homogeneity. Figure 2a also shows that two groups are clearly distinguished. The first group consists of V1nn,

V2bb and V3pp variables **(**Cluster A) and represents control group pre-service teachers' efficiency in problem solving related to all types of the questioned distributions. The second group includes variables V4pp, V6nb and V5bb variables **(**Cluster B) and concerns the case of solving the problem tasks connected with Normal, Binomial and Poisson distribution. The strongest similarity occurs between variables V1nn and V2bb (0.999018) (see Figure 2a). Also, the answers of Control group' to task V3pp come along with variables V1nn and V2bb from Cluster A, indicating their' coherence in dealing with the corresponding tasks, irrespectively of distribution types. In addition, the similarity of Cluster A is also strong (0.967745).

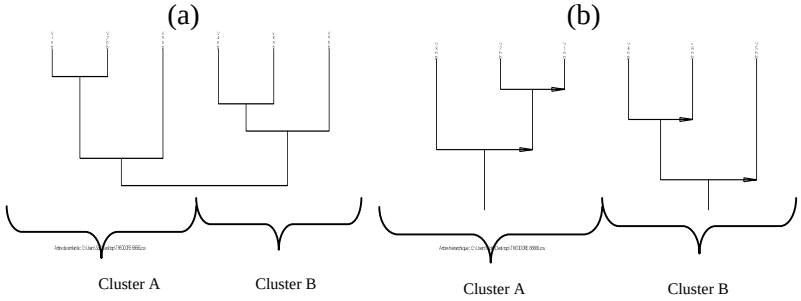

Figure 2. (a) Similarity diagram and (b) hierarchical tree of control group

In Cluster B, the strongest similarity occurs between variables V4pp and V6nb (0.99814) and suggests that pre-service teachers of the control group employed similar processes to construct a problem solving strategy with regard to different kinds of tasks. Also, the similarity of the second group is very strong too (0.981403). The hierarchical tree (Figure 2b) of the control group's answers accords with the similarity diagram as it shown in Figure 2a.

CONCLUSIONS

In this paper the problem of pre-service teachers' approaches in solving tasks in probability distributions is discussed. Two groups of students took part in the study. The control group participates in teacher-centred teaching environment. The experimental group participates in student-centred teaching environment. Experimental group students were allowed to become involved to creation of their own task along with academic demands. Those students got a more meaningful learning and achieved higher performance. In addition the estimation of the Poisson distribution parameter λ troubled the majority of the control group because it presented low percentages. Further investigation of this issue will take place in a future work, with the use of a more extended analysis. Besides, longitudinal investigations might reveal new insights about how the effectiveness in using the types of projects grows.

REFERENCES

Biajone J. (2006). Promoting positive attitudes toward statistics in pedagogy students through projects work. *ICOTS 7.* Online: www.stat.auckland.ac.nz/~iase/publications/17/C408.pdf.

César, M., & Oliveira, I. (2005). The curriculum as a mediating tool for inclusive participation: A case study in a Portuguese multicultural school. *European Journal of Psychology of Education, 20*(1), 29-43.

Chadjipadelis, T., & Gastaris, P. (1995). Difficulties of understanding and misconceptions in Probabilities and in Statistics. *Eucledes C'* 43, 35–68.

Chadjipadelis T. (2003a). Children's walk in probability and statistics from childhood to youth. In *Proc. 19th Panhellenic Conference on Mathematical Education, Veria* (pp. 27-40).

Chadjipadelis, T. (2003b). Data, methods and results. In *Proc. 16th Panhellenic Conference on Statistics, Kavala* (pp. 37-44).

Guimarães, H. M., Silva, A., Ponte, J. P., Santos, L., Abrantes, M., & Abrantes, P. (Eds.) (2005). *Paulo Abrantes: Intervenções em Educação Matemática.* Lisboa: APM.

Ponte, J. P. (1990). *O Computador, Um Instrumento da Educação*. Lisboa: Texto Editora.

Zeleke, A. Lee, C., & Daniels, J. (2006). Developing project based on students' data in introductory statistics. *ICOTS7*. Online: www.stat.auckland.ac.nz/~iase/publications/ 17/8C1_ZELE.pdf.