COMPUTING PROBABILITIES FROM TWO WAY TABLES: AN EXPLORATORY STUDY WITH FUTURE TEACHERS

Assumpta Estrada Roca
University of Lleida, Spain
Carmen Díaz Batanero
University of Granada, Spain
aestrada@matematica.udl.es

Being able of correctly read and interpret two ways tables is a basic component of statistical literacy for every citizen. Therefore, future teachers who will be responsible to teach statistics to children at school level should acquire these abilities along their training. However, this capacity is taken for granted in Spain and its teaching is not usually included in the curriculum for training teachers. In this study present the results of a small exploratory study that describe the future teachers' semiotic conflicts in solving elementary probability problems when data re given in a two-way table.

INTRODUCTION

In the past years Statistics is increasingly taking part in the primary school mathematics curriculum in most countries and voices claming for statistical literacy, statistical thinking and basic knowledge for every citizen are stronger, since they are essential tools in the information society (Gal, 2002). These claims have been, however not answered in Spain neither at level of children education in the schools, nor in the training of teachers at the Faculties of Education. As suggested by Silva *et al.* (1999), future teachers should be motivated and acquire sufficient skills to learn the statistical techniques needed to teach the statistical contents included in the curricula and to evaluate the quality of their professional work.

However, in practice few teachers teach Statistics in schools. In Gattuso and Panone (2002) interviews, teachers themselves recognise this is the part of Mathematics they reduce more, because they think there should prepare their students in more important topics for the final evaluations. In addition, according to Estrada (2005), the teachers' specific formation in Statistics and its Didactics is practically nonexistent and they generally consider statistics as application of routine formulas, and not as an multidisciplinary tool indispensable for their academic and professional life. All these reasons suggest us the need to reinforce the specific and didactic preparation of statistics teachers, and also the need of assessing the teachers' difficulties and errors in learning the topic.

Our study is focused on two way tables, because statistical literacy include the capacity to interpret and critically evaluate statistical information and make decisions based on random phenomena that people find in diverse personal and professional contexts (Gal, 2002) and this information basically appears in statistical graphs and tables. Our aim is to show that a statistical two-way table is a complex semiotic object and to assess teachers' ability to read and compute probabilities from it. We analyse the responses to elementary questions about simple, joint and conditional probabilities from these tables in a sample of 65 future teachers from the University of Lérida (Spain). We base on ideas taken from Godino (2003), who considers that mathematical activity is essentially relational, and mathematical objects are put in relation in this activity by means of semiotic functions. Using this theory, we carry out a detailed analysis of the way these objects are related by teachers when solving the task of reading a table, with the purpose of identifying semiotic conflicts or disagreements between the meaning attributed to a same expression by two students or by a student and the teacher.

PREVIOUS RESEARCH

A two-way or contingency table serves to present in a summarised way the frequency distribution in a population or sample that was classified according two statistical variables. The simplest way, where each variable has only two categories is presented in Table 1. Usually two-way tables are more complex, with several rows or columns and include several types of totals, so that the semiotic conflicts we identify in this paper might increase in more complex tables.

Table 1: Typical format for a 2x2 contingency table

	A	*No A*	*Total*
B	*a*	*B*	*a+b*
no B	*c*	*D*	*c+d*
Total	*a+c*	*B+d*	*a+b+c+d*

Didactic research on contingency tables has focused on students' capacity to assess association between the variables A and B from the data presented in a 2×2 contingency table. Some examples are research by Smedslund, 1963, Jenkins and Ward (1965) and Estepa (1993), all of them basing on the pioneer study by Inhelder and Piaget (1955), who considered the understanding of association as the final step in developing the idea of probability, and that understanding association has as prerequisites the comprehension of proportionality, probability, and combinatorics. Consequently, they only studied reasoning about association with children in their formal operation stage. None of this research questioned the students' capacity to read the data in Table 1, which is a complex semiotic object. Data in cells *a, b, c, d* refer to joint absolute frequencies, each of them for a double condition (values of row and column). Even in the easy case of a 2x2 table these cells are non-equivalent. A high frequency in cells *a* (presence of character *A;* presence of character *B*) and *d* (absence of *A*, absence of *B*) would indicate a positive association between the variables; and the other two cells a negative association.

This is not easily perceived. Moreover, from a given cell we can deduce joint as well as row and column conditional relative frequencies. All these mathematics objects co-exist and may be confused by students when interpreting the data table, in either interpreting association between the variables *A* and *B* or in computing probabilities. For example in Batanero *et al*. (1996) around 20% students confused column and row conditional relative frequencies. It is also possible that this same confusion appear in computing probability from the table. We also should take into account Falk's (1986) remark that many students do not adequately discriminate the two different conditional probability *P*(*A/B*) and *P*(*B/A*). The author speak of the *fallacy of transposed conditional*. Pollatsek, Well, Konold and Hardiman (1987) suggest that performance in conditional probability problems depend on the way the statement is posed- For example Einhorn and Hogarth (1986) observed that some students misinterpreted the conjunction "and," and confuse joint and conditional probability. This mistake also appeared in Ojeda' s (1995) research, where about half students misinterpreted conjunction as conditioning. Finally Huerta and Lonjedo (2005) showed the influence of data format on solving simple conditional probability problems.

METHOD

The sample in the study were 65 future teachers at the Faculty of Education, University of Lerida, Spain. We gave them item 1. which takes part of a wider questionnaire that assesses understanding of conditional probability (Diaz, 2005). These students had attended a course in Mathematics and its Didactics, including a Statistics unit before taking the test.

Item 1: In a medical centre 780 people have been observed. The following results were obtained:

	Younger or equal to 55	Older than 55	Total
The person had a heart stroke	29	75	104
The person never had a heart stroke	401	275	676
Total	430	350	780

Suppose we pick one of these people at random:

a) Which is the probability that the person had a heart stroke?
b) Which is the probability of being older than 55 and at the same time having had a heart stroke?
c) If the person selected is older than 55, which is the probability that the person had a heart stroke?

d) If the person had a heart stroke, which is the probability of being older than 55?

Once the answers were gathered we categorized them, taking into account the concepts used by the students. Responses are presented in Table 2, where we use the following abbreviations A="A person had a heart stroke;" no A="The person never had a heart stroke;" B = "the person is older than 55 years," no B = "the person is younger or equal to 55." Correct responses are shadowed. We can observe that, although the majority of future teachers correctly compute simple probability, there are around 40- 50% of errors in computing compound and conditional probabilities.

Table 2: Frequency (and percentage) of responses to the four tasks

Teacher's answer	Task requested			
	P(A)	P(A∩B))	P(A/B)	P(B/A)
P (A)	49 (75.3)		1	
P(B)	1			
P(A∩B)	1	34 (52.0)		1
P(A/B)		4	36 (56.0)	9
P(B/A)		8	5	34 (52.0)
P(no A∩B)		2	1	1
Assuming independence		1		
P(A/ no B)	1		1	
P(noA/ no B)			1	
Computing possible cases		1	1	
Other errors		1		
Blank	13	15	18	20

We remark the big number of students who did not provide any solution; they justified in their responses that "*I cannot remember,*" "*It never was explained to me,*" "*There is time since I studied this,*" "*I knew, but I am forgotten.*" In the task of computing simple probability the only error was confusing simple and compound probability. There are a greater variety of errors in the other tasks, in particular confusion between compound and conditional probability. At the same time we found other mistakes which have not been described in previous research, such as confusing an event and its complementary, confusing probabilities with possible cases (absolute frequencies) or assuming independence in the data. In Table 3 we classify these errors according the different semiotic conflicts between the mathematical objects involved in the four tasks given to the students, where some responses included more than one semiotic conflict.

Table 3: Types of semiotic conflict identified in the teachers' responses

	N=65
Confusing a conditional probability and its inverse	14
Confusing conditional and joint probability	13
Confusing an event and its complementary	7
Assuming independence	1
Confusing simple probability with either conditional or joint probability	3
Computing possible cases (absolute frequencies)	2
Other errors	2
Total number of conflicts	42

CONCLUSIONS

Our results suggest that reading a two-way table and computing probabilities from them is not easy for future teachers. Our study provides some information in this sense and complement previous research by Estepa (1993) and Batanero *et al*. (1996) about future teachers understanding of elementary statistical concepts.

We agree with Falk (1986) that the everyday language we use to state a conditional probability problem lack enough precision and is therefore ambiguous. However a future teacher should master both the concept and the language used in teaching, particularly those that take today part of statistical literacy, which is important for their students, and they should transmit them.

Still, the numerous semiotic conflicts and the number of students which were unable to provide a response should make reflect the teachers trainers and suggest them the need to reform and improve the statistics education these future teachers are receiving during their training in the schools of education.

REFERENCES

Batanero, C., Estepa, A., Godino, J. D. and Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education,* 27(2), 151-169.

Díaz, C. and de la Fuente, I. (2005). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística (Reasoning on conditional probability and implications for teaching of statistics). *Epsilon,* 59, 245-260.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic and Tversky (Eds.), *Judgement Under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

Estepa, A. (1993). *Concepciones Iniciales Sobre la Asociación Estadística y su Evolución Como Consecuencia de una Enseñanza Basada en el Uso de Ordenadores*. Thesis Doctoral. Granada: Departamento de Didáctica de la matemática, Universidad de Granada.

Estrada, A. (2005). A structural study of future teachers' attitudes towards statistics. Paper presented at the CERME IV Conference (European Research in Mathematics Education Conference), Sant Feliu de Guisols, Gerona, Spain.

Einhorn, H. J. and Hogart, R.M. (1986). Judging probable cause. *Psychological Bulletin,* 99, 3-19.

Falk, R. (1986). Conditional Probabilities: Insights and difficulties. In R. Davidson and J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics,* (pp. 292 – 297). Victoria, Canada: International Statistical Institute.

Gal, I. (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review,* 70(1), 1-25.

Gattuso, L. and Pannone, M. A. (2002). Teacher's training in a statistics teaching experiment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics,* Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Godino, J. D. (2003). *Teoría de las Funciones Semióticas. Un Enfoque Ontológico-Semiótico de la Cognición e Instrucción Matemática* (Theory of semiotic functions. An ontological and semiotic approach to mathematics cognition). Granada: Departamento de Didáctica de la Matemática, Universidad de Granada, http://www.ugr.es/local/jgodino/indice_tfs.htm.

Inhelder, B. and Piaget, J. (1955). *De la Logique de L'enfant a la Logique de L'adolescent* (From the logic of children to the logic of adolescent). Paris: P.U.F.

Huerta, P. and Lonjedo, M. A. (2005). The nature of the quantities in a conditional probability problem. Its influence in the problem resolution. Paper presented at IV CERME, European Research in Mathematics Education, Sant- Feliu de Guitxols, Gerona, Spain.

Jenkins, H. M. and Ward, W. C. (1965). Judgement of contingency between responses and outcomes. *Psychological Monographs,* 79.

Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO,* 5, 37-55.

Pollatsek, A., Well, A. D., Konold, C. and Hardiman, P. (1987). Understanding conditional probabilities. *Organitation, Behavior and Human Decision Processes*, 40, 255 – 269.

Silva, C. B., da. Cazorla, I. M. and Brito, M. R. F. (1999). Concepçoes e attitudes em relaçao a estatistica (Conceptions and attitudes regarding statistics). From *Conferência Internacional: Experiências e Perspectivas do Ensino da Estatística*, (pp. 18-29), Florianópolis

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology,* 4, 165-173.