STUDENTS' INTUITIVE STRATEGIES IN JUDGING ASSOCIATION WHEN COMPARING TWO SAMPLES

Antonio ESTEPA
University of Jaén, Spain
Carmen BATANERO
University of Granada, Spain
F. T. SÁNCHEZ
University of Jaén, Spain
(Received January 5, 1998)

Abstract

This paper is an experimental study on students' strategies and association judgements when faced with the comparison of a numerical variable in two different samples (comparison of two samples). The strategies are classified from a mathematical point of view, which allows us to identify theorems in action and two types of misconceptions about association.

STUDENTS' INTUITIVE STRATEGIES IN JUDGING ASSOCIATION WHEN COMPARING TWO SAMPLES

Statistical association is a relevant area in mathematics education, because it extends functional dependence and has significant connections with probability and proportional reasoning. Studying association among variables is fundamental in experimental research in different sciences (e.g., biology, economics, medicine, education) to find causal explanations, that permit us to understand our environment. However, the existence of association between two variables does not necessarily imply a causal relationship, because it is possible to find a high coefficient of association among variables when there is no causal link (spurious correlation), due to the concurrence of different factors affecting the response variable.

Beyond these epistemological problems, psychological research has shown that judging association is not an intuitive ability. Adults sometimes use only part of the available data to judge association and base their judgment on their previous beliefs about the type of association that ought to exist between the variables studied rather than on the empirical data. Chapman & Chapman (1969) have described this phenomenon as "illusory correlation" because people maintain their beliefs in spite of evidence of data. Jennings, Amabile and Ross (1982), Wright and Murphy (1984), and Alloy and Tabachnik (1984), among others, have studied the effect that previous theories, about the context of a problem, have on judging association. The general conclusion is that when there is agreement between data and previous expectations people increase their confidence in the attribution of co-variation. On the other hand, when data does not coincide with these expectations, there is a cognitive conflict and the accuracy of the perception of co-variation depends on the relative strength of the two sources of information.

Determining this preconception is highly relevant for the teaching of association. As Confrey (1990) has pointed out, the need of research on students' conceptions lies in the fact that sometimes these conceptions differ in fundamental ways from the scientific concepts that we try to teach. Furthermore, students' conceptions are resistant to change in spite of instruction. Although many students in our study demonstrated correct or partially correct judgements and solutions strategies, the misconceptions and incorrect strategies presented in this article indicate a gap between the meaning of association that we try to teach, and the subjective meaning that students might attribute to this concept (Godino & Batanero, 1994).

Background

This study continues our previous research on statistical association (Batanero et al., 1996, Estepa & Batanero 1996, Batanero et al., 1997), where we studied students' preconceptions, association judgements and solving strategies in contingency tables and scatter plots and their evolution after a teaching experiment based on the use of computers. In the aforementioned papers the following misconceptions concerning statistical association: were described:

a) Determinist conception of association: Some students do not admit exceptions to the existence of a relationship between the variables. They expect a correspondence that assigns only one value in the response variable for each value of the explanatory variable. When this is not so, they consider there is no association between the variables. For example, some students argue that cells out of the diagonal in a 2x2 contingency table ought to have zero frequency in case of association. In scatter plots they expect no spread in the plot and sometimes they even look for an algebraic equation relating the two variables.

b) Local conception of association: Students often form their judgement using only part of the data provided in the association problem. If this partial information serves to confirm a given type of association, they would adopt this type of association in their answer. For example, in contingency tables, students sometimes base their judgement on only one conditional distribution or even on only one cell, usually the cell for which the frequency is maximum. These students failed to grasp what in opinion of Konold et al., (1997) is the essence of the statistical perspective: attending to features of aggregates as opposed to features of individuals, as distributions and their parameters are features belonging not to any individual element, but to the aggregate which they comprise.

c) Unidirectional conception of association: Some students perceive the association only when their sign is positive, considering the inverse association independence and in other cases they have trouble with the interpretation of negative correlation coefficient.

d) Causal conception of association: Some students only consider the association between the variables if this could be attributed to a causal relationship between them. This misconception was persistent in our teaching experiment (Batanero et al., 1997), and the theoretical study of the relationships between correlation and causality during the teaching was insufficient to change the students' conception.

In this article we report on an experimental study whose aim was to find out whether some of the aforesaid described preconceptions concerning association were found when students compare two samples on a numerical variable. Some authors, such as Konold et al. (1997) have explored students' ability to determine whether there is a difference between groups. Although judging group differences is formally comparable to judging association, these two types of judgements probably describe different cognitive tasks in opinion of Konold et al. (1997) who suggests that the methods used by students in these problems would depend on the way the question is posed and the reason for doing the comparison.

Our data was taken from a sample of pre-university students, through their written responses to a questionnaire including two problems. After describing the problems, we, firstly, discuss the type of association perceived by the students in the two different tasks. We, secondly, analyse the students' solving strategies, and present an original classification of these strategies, from a mathematical point of view, in which we identify concepts and theorems in action as described by Vergnaud (1982). Finally, we describe the different misconceptions found concerning statistical association.

METHOD

Sample

The sample consisted of 213 students in their last year of secondary school (17-18 year-old students) attending three different high schools in the city of Jaen (eight different student groups were used). It is at this level that the topic of association is introduced in the Spanish curriculum. About half the students (113) were males and half where females (100). The questionnaire was given to the students before the teaching of association began, so that this study may be considered as research into students' preconceptions concerning statistical association.

One hundred and twenty four students had followed a scientifically-oriented curriculum and eighty nine students had followed a humanities-oriented curriculum. Although both groups of students had studied mathematics throughout their previous studies, in the first group a stronger emphasis was given to the subject and they had more opportunities to apply mathematics in other scientific subjects, such as physics or chemistry. Concerning statistics, both groups were taught frequency distribution, graphical representation, measures of location, and spread of univariate statistical variables. They had also been introduced to probability in the first year of secondary school (when they were 14-15 years old). This study has a quasi-experimental character, because of the non-random nature of the samples of students and problems.

Questionnaire

The students were given the two questions concerning the comparison of two samples that are shown in figures 1 and 2. In the first item the samples are related whereas they are independent in the second item. The t- test of mean differences between the two samples was computed for each problem, giving a p-value of p= 0.00092, t= 3.3 in the first item and p= 0.1510, t= 1.5 in the second one.

Data analysis

In each item we analysed the type of association perceived by the students (direct association, inverse association, or independence). In addition, the students' written responses to the questionnaire were categorised and, after several successive revisions, a scheme was developed for classifying students' solution strategies based on Vergnaud's (1982) notion of *theorem in action*.

Figure 1
Item 1. The following data were obtained when measuring the blood pressure to a group of 10 women, before and after applying medical treatment:

	Mrs. A	Mrs. B	Mrs. C	Mrs. D	Mrs. E	Mrs. F	Mrs. G	Mrs. H	Mrs. I	Mrs J
Blood pressure before treatment	115	112	107	119	115	138	126	105	104	115
Blood pressure after treatment	128	115	106	128	122	145	132	109	102	117

Using the information contained in this table, do you think that the blood pressure in this sample depends on the time of measurement (before or after the treatment)?

Explain your answer.

Figure 2
Item 2. The following data were obtained when measuring the sugar level in the blood of male and female school children:

Pupil	A	B	C	D	E	F	G	I	J	K	L	M	N	O	P	Q	R	S	T	U
Gender	M	M	M	M	M	M	M	M	M	M	F	F	F	F	F	F	F	F	F	F
Sugar level	9	0	9	8	6	7	4	9	8	9	6	0	7	0	8	3	6	7	7	3

M = male, F = female.

Using this information, do you think that the sugar level in the blood in this sample depends on the sex?

Explain your answer.

A pilot study with an additional sample of 51 students was used to check the reliability and the coding system for the students' answers. Since factor analysis of students' answers had a multidimensional structure, a generalizability study (Brennan, 1983) was done, instead of computing an internal consistency reliability index. We obtained a generalizability index, G= 0.86, as a measure of the possibility of extending our conclusions to the hypothetical item population and another index, G= 0.94, of generalizability to the subjects' population. This pilot sample was also used to perform a qualitative study of the interrelation between strategy and the type of association perceived by the students, in order to characterise students' misconceptions concerning statistical association. The results of all of these analyses are presented in the following sections.

DISCUSSION

In table 1 frequencies and percentages of answers concerning the existence of association between the two variables in each item are shown. In both cases the majority of students gave a correct response, so that judging the association in this type of problem was not a difficult task for our students.

Table 1. Frequencies and percentages of association judgements in the comparison of two samples.

Item	Association	Independence	No answer
1	*178 (83.6)	20 (9.4)	15 (7.0)
2	*157 (73.7)	44 (20.7)	12 (5.6)

*correct answer

However, the correct judgement of the existence of association between the two samples is not enough to determine the preconceptions that the students have about this topic, since, as different research has shown, we can get a correct solution for a given problem using an incorrect procedure. From the perspective of Mathematics Education, both a correct procedure and a correct Judgment of association are needed to correctly understand the idea of association.

Making a judgement concerning the association of a numerical variable with regards to a dichotomic qualitative variable (as in the comparison of two samples) is a task that must be solved in the numerical frame (Douady, 1986). Furthermore, as the response variable is quantitative, we can compare its different statistics, such as averages and ranges, in the two samples, to solve the problem. Graphical representations of the two distributions could also be used.

In studying the strategies, we noticed that some students proved their correct intuitive conceptions as they solved the problem by using normative strategies. But the concept of association is not simple and some students might have correct intuitive conceptions concerning some properties of association mixed with other incorrect conceptions. This would lead them to choose an incorrect or partially correct strategy. All these conceptions need to be identified in order to plan adequate instruction.

In order to identify which of the properties linked to the concept of association were intuitively used correctly or incorrectly by our students, we developed a classification of students' strategies according to the mathematical concepts involved and theorems in action implied in the different procedures used and the type of errors.

Such a classification is, in Vegnaud's (1982) opinion, essential for didactical analysis. In the following paragraphs we describe this classification.

Correct Strategies

1. *Comparing means:* Some students compared the two mean samples to decide the existence of association between the variables, such as Student 43 in the second item: *"Because when you add the sugar levels in all the males you obtain 69, that divided by 10 is 6.9, while the sum of sugar levels in females is 47 that divided by 10 is only 4.7. Therefore, there is dependence on the sex, as sugar level is higher in males". The student might implicitly be using the following theorem in action: "When there is a sufficiently high difference between the means of the conditional distributions of a quantitative variable with regards to the values of another qualitative variable, there is an association between the two variables" (T1).* This is an example of a correct answer for the institutional relation introduced to the student, which does not include the usual procedures for testing the equality of means, where it is necessary to consider the variability of the two data sets, the sample sizes and the significance level, in addition to the mean values.

2. *Comparing totals:* Some students compared the totals of the response variable values in each sample: *"When adding all the sugar levels in males we obtain a grater number than when adding all the sugar levels in females" (Student 31).* They might be using the fact that dependence would imply the variation of the sum of values in the two distributions, which is a modification of the previous theorem in action T1.

3. *Comparing percentages:* In Item 1, some pupils compared the two values of blood pressure in each woman and based their Judgment on the proportions or percentage of cases where the pressure increased or decreased. *"We can observe that the pressure decrease in 20% of women and increases in the remaining 80%", and, therefore, the treatment increases the pressure"* (Student 8).

In this case, a grouping of women in two categories –according to whether the pressure increased or decreased– was produced. These students related the solution to a probability problem and used Laplace's rule, considering that if there were no dependence, these categories should have the same frequency. This procedure is similar to that used in parametric methods based on rank comparisons in a series. Therefore, the students might be using the following theorem in action: *"If we take common values in both samples, the percentages of cases in which the value of the variable increases or decreases would be the same in case of independence" (T2).*

4. *Comparing the two distributions:* Other students compared the complete distribu-

tions of frequencies in both samples, as in the following example: "*When we compare males and females, the latter have got a lower sugar level. That is proved by the fact that two girls have 0, while there is only a boy with 0; There are 2 girls with 3, while there isn't a boy with that pressure; there is only 1 boy with value 4. There are 2 girls with value 6 and only one boy with 6. There are 3 girls with 7 and only 1 boy with 7. There are 2 boys with 8 and only 1 girl with 8. There are 4 boys with 9 and no girl had 9*" (Student 12).. They might be using the following theorem in action: "*There is independence when both samples have the same frequency distributions" (T3).*

Partially Correct Strategies

5. *Comparing the values of the response variable for each case in related samples.* These students compared the values corresponding to the same subject in both samples. It is very similar to strategy 3, though here proportions or percentages were not used, and, therefore, we consider this strategy to be only partially correct. Some students failed in their association judgement when using this strategy, because they expected the difference to have the same sign in every case, which indicates a *functional or determinist* conception of the association: "*I have compared them and I found out that after the treatment blood pressure decreased in some women, whereas in other cases blood pressure increased. There is no dependence*" (Student 30).

6. *Taking into account exceptional cases.* Sometimes, a correct Judgment was obtained by considering the cases where the general rule was not followed (exceptional cases). We consider the strategy partially correct, because it was used mainly to indicate an association. As the students did not use proportions or percentages and based their judgement in only part of the distribution, a local conception can be deduced in these students. "*Blood pressure increases in all the women except in Mrs C, so you can say that blood pressure depends on treatment*" (Student 29).

7. *Finding out the differences.* The students computed the differences between the values of the response variable in every subject, in related samples, basing the association judgement on these differences: "*I have computed the differences in pressure before and after treatment in every woman and most of these differences are positive*" (Student. 5). These pupils might have used the theorem in action: "*In order to find the association between a continuous and a dichotomic variable the difference in the corresponding values of the continuous variable must be large enough" (T4).*

8. *Global comparison.* When the association judgement was carried out by means of a qualitative global comparison of both samples, without specifying which statistics were used. For example, "*In the data we can see that the males have got a higher sugar*

level with regards to the females" (Student 29).

Incorrect Strategies

As indicated by Nisbett and Ross (1980), subjects do not usually employ a normative strategy in judging association without being given specific instructions because these strategies are highly sophisticated. We found the following incorrect strategies in our students:

9. Expecting similar values. Sometimes similar values in all the cases for each sample were expected. As that did not happen, because of the data randomness, they concluded that the blood pressure was independent of the treatment, showing a *determinist conception* of the association. " *You can see a lot of variety in the table; some women with a high pressure have increased and other have decreased and vice-versa"* (Student 51).

10. Comparing highest and lowest values of both distributions, and basing the association judgement on them, which indicates a local conception of the association. *"It does not depend on sex as some males have the highest sugar level and others have the lowest."* (Student 208).

11. Comparing ranges. Some students compared the range of both samples, such as student 66: *"Because you can see that females got the lowest number (0) as well as males. However, they do not get the highest number (9) as males do".*

12. Assessing coincidences. Some students justified independence because of the coincidence of some cases, just as the student who answers: *"Because some males have the same sugar level as some females, such as the pupils F and N who have got the same sugar level of 7 in spite of being different, male and female"* (Student 133). Konold et al. (1997) indicates that comparing two or more individuals in respect to some attribute is a frequently used non- statistical comparison. In this example, we can add a *determinist conception* of association.

13. Basing on previous theories. Some students based their association judgement on their knowledge about the context (previous theories) instead of using the data given in the problem like this student's answer: *"It depends on whether you are examined by a male or female doctor when you don't feel very well. At that moment you might get a bit excited and that's the reason for the increase of blood pressure".* Therefore, the phenomenon of illusory correlation, might be acting in these students (Chapman and Chapman, 1969).

14. Others. In this category other kinds of procedures and strategies have been included, such as students who compared the means, which was computed without taking into account null values (phenomenon that have been described by Mevarech, 1983), or using the values of the variables as if they were percentages, confusion very common in students according to Hawkins et al., (1992).

Table 2. Frequencies and percentages of students' strategies in the comparison of two samples.

STRATEGIES	ITEM 1	ITEM 2
Correct		
1. Comparing means	2 (0.9)	26 (12.2)
2. Comparing totals	1 (0.5)	35 (16.4)
3. Comparing percentages	41 (19.3)	1 (0.5)
4. Comparing the distributions		34 (16.0)
Partially correct strategies		
5. Comparing all cases	73 (34.3)	3 (1.4)
6. Assessing exceptional cases	63 (29.6)	
7. Finding out differences	6 (2.8)	
8. Global comparison		64 (30.0)
Incorrect strategies		
9. Expecting similar values	3 (1.4)	1 (0.5)
10. Comparing highest and lowest values		6 (2.8)
11. Comparing ranges		3 (1.4)
12. Assessing coincidences	1 (0.5)	3 (1.4)
13. Expressing previous theories	1 (0.5)	8 (3.4)
14. Others	8 (3.8)	14 (6.6)
15. No answer	14 (6.6)	15 (7.0)

Correct and Partially Correct Conceptions

In Table 2 we present the frequencies and the percentages of these strategies in the students' answers. Correct strategies were more frequent in independent samples (item 2), while we can find a higher percentage of partially correct strategies in related samples (item 1). It is worthwhile mentioning that some strategies, such as strategies 3,5,6,7, and 9, were characteristic of related samples, while strategies 1, 2, 4, 8, 10, 11, 12, and 13 appear mainly associated with independent samples. Therefore, even though the students had not formally studied the difference between independent and related samples, they were able to distinguish them when comparing the samples for study in the existence of association.

Students who compared the means in independent samples, conceived the mean as a value representative of the data (Pollatsek and cols, 1981, Mevarech,1983). Leon and

Zawojewski (1991) consider that this indicates a relational knowledge of the mean (Skemp, 1978), consisting in having suitable schemes or enough conceptual structures available to solve a wide range of problems where the concept is necessary. The same can be said of students comparing totals (16.4%), since these totals are values representing the set of data. Moreover, choosing between various summaries in a concrete context requires knowledge on relations between distributions and summaries and functional interpretations of summaries (Biehler, 1997).

The strategy based on comparing percentages (strategy 3) transformed the original problem into a contingency table 2x 2, where it was possible to use a probability comparison. Therefore, these students applied general heuristics problem-solving.

One of the most frequent questions the researcher faces when two samples are compared, is whether these samples come from the same population or not. This is what the students who compared distributions did intuitively (strategy 4). From our results we can infer that the strategy of comparing corresponding values in related samples can be found, at different levels of elaboration in the 87.9 per cent of students (Strategies 5, 6 and 7).

When the data are globally compared (strategy 8) the student is implicitly accepting that all the data are relevant, which is a basic step in the association judgement (Crocker, 1981).

Misconceptions Found

The *determinist conception* of association appeared when the student expected the variation in the data to always have the same sign (25 (11.7%) of students using strategy 5) or when he/she expected similar values of the dependent variable in each sample (strategy 9).

The *local conception* can be noticed when the student used isolated cases to decide his/her association judgement; for example, when using exceptional cases (strategy 6) or comparing extremes in both samples (strategy 10). These students have not adopted the propensity perspective described by Konold et al. (1997) and make sense of statistics by using more familiar, but inappropriate comparison schemes.

CONCLUSIONS

Our results show a mixture of correct and incorrect preconceptions and strategies in the Judgment of association when comparing samples. We also deduce that comparing samples to judge association is a highly skilled activity even at an exploratory level, requiring a varied statistical knowledge. It also implies selecting the best statistical procedure, adequate interpretation of results and relating them to the research questions concept and using it in problem solving. All our findings show the complexity of a topic simple in appearance. Due to this complexity and the relevance of understanding the idea of association, we conclude that there is a need to extend this research and reinforce the teaching of this topic at university level and in the final years of secondary school. This will require informing teachers about the mathematical, psychological, and educational aspects of the topic, including information about students' preconceptions as set out in this article.

Acknowledgment: This report forms part of the Projects PB96-1411 and PB97-0851, granted by Secretaria de Estado de Universidades, Investigacion y Desarrollo (M.E.C.). Madrid (Spain).

REFERENCES

Alloy, L. B. & Tabachnik, N., (1984), Assessment of covariation in humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, 91, 112-149.

Batanero, M.C., Estepa, A., Godino, J.D., & Green, D., (1996), Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27 (2), 2-19.

Batanero, C., Estepa, A., & Godino, J. D., (1997), Evolution of students' understanding of statistical association in a computer-based teaching environment. In J. Garfield & G. Burrill (Eds), *Research on the Role of Technology in Teaching and learning Statistics. 1996 IASE Round Table Conference Papers* (pp. 191-205), International Statistical Institute, Voorburg.

Biehler, R., (1997), Students' difficulties in practising computer supported data analysis. Some hypothetical generalisations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds), *Research on the Role of Technology in Teaching and learning Statistics. 1996 IASE Round Table Conference Papers.* (pp. 169-

190), International Statistical Institute, Voorburg.

Brennan, R. L., (1983). *Elements of Generalizability Theory*. ACT Publications, Iowa.

Chapman, L.J. y Chapman, J.P., (1969). Illusory correlation as an obstacle to the use of valid Psychodiagnostic signs. *Journal of Abnormal Psychology*. 74, 271-280.

Confrey, J., (1990). A review of the research on students conceptions in Mathematics, Science, and Programming. *Review of Research in Education*, 16, 3-55.

Crocker, J., (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*. 90, 2, 272-292.

Douady, R., (1986). Le jeu de Cadres et dialectique outil-objet dans l'enseignement des Mathematiques. [Set of patterns and tool-object in teaching mathematics]. *Recherches en Didactique des mathematiques*, 7.2, 5-31.

Estepa, A. & Batanero, M. C., (1996). Judgements of correlation in scatter plots: students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 21-41.

Godino, J. D. & Batanero, M. C., (1994). Significado institucional y personal de los objetos matematicos. [Institutional and personal meaning of the mathematical concepts.] *Recherches en Didactiques des Mathematiques*, 14, 3, 325-355.

Hawkins, A., Jolliffe, F. y Glickman, L., (1992). *Teaching statistical concepts*. Longman, Nueva York.

Jennings, D. L., Amabile, M. T. & Ross, L., (1982). Informal covariation assessment: Data-based versus theory-based judgements. In D. Kahneman, P. Slovic & A. Tversky (Eds), *Judgment under uncertainty: Heuristics and biases*, (pp. 211-230). Cambridge University Press, New York.

Konold, C., Pollatsek, A., Well, A. & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds), *Research on the Role of Technology in Teaching and learning Statistics. 1996 IASE Round Table Conference Papers* (pp. 151-167) , International Statistical Institute, Voorburg.

Leon, M.R. and Zawojewski, J.S., (1991), Use of the arithmetic mean: an investigation of four properties issues and preliminary results. En D.Vere-Jones, (Ed.), *Proceedings of the Third International Conference on Teaching Statistics*, Vol. 1, (pp. 302-306). Otago University Press, New Zeland.

Mevarech, Z. R., (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics*, 14, 415-429.

Nisbett, R. & Ross, L., (1980), *Human inference: strategies and shortcoming of social judgements*. Prentice Hall, New Jersey.

Pollatsek, A., Lima, S. and Well, A.D., (1981), Concept or computation: Students' understandings of the mean. *Educational Studies in Mathematics*, 12, 191-204.

Skemp, R., (1978), Relational understanding and instrumental understanding. *Arithmetic Teacher*, Nov., 9-15

Vergnaud, G., (1982), Cognitive and developmental Psychology and research in Mathematics Education: some theoretical and methodological issues. *For the learning of mathematics*, 3, 2, 31-41.

Wright, J. C. & Murphy, G. L., (1984), The utility of theories in intuitive statistics: the robustness of theory-based judgements. *Journal of Experimental Psychology*, 113(2), 301-322.

Antonio Estepa *University of Jaén*
Faculty of Humanities and of Sciences of the Education
Department of Didactic of the Sciences
Campus of the Lagunillas, s/n. Building D-2
23071 Jaén
Spain
e-mail: aestepa@ujaen.es

Carmen Batanero *University of Granada*
Faculty of Sciences of the Education
Department of Mathematics Education
Campus of Cartuja
18071 Granada
Spain
e-mail: batanero@goliat.ugr.es

Francisco T. Sánchez *University of Jaén*
School Polytechnic
Department of Mathematics
Avenue of Madrid, s/n.
23071 Jaén
Spain