JUDGMENTS OF CORRELATION IN SCATTER PLOTS: STUDENTS' INTUITIVE STRATEGIES AND PRECONCEPTIONS

Antonio ESTEPA
University of Jaén, Spain
Carmen BATANERO
University of Granada, Spain
(Received May 16, 1995)

Abstract

In this paper we describe an experimental study of 213 pre-university students' strategies when assessing correlation in scatter plots. We present an original classification of students' strategies from a mathematical point of view, which allows us to determine concepts and theorems in action. Correspondence analysis is used to show the effect of task variables of the items on students' strategies and, finally, we describe three different students' misconceptions concerning statistical association.

INTRODUCTION

The concepts of correlation and regression, and, in general of association, extend functional dependence and are fundamental for many statistical methods that are used in making predictions and taking decisions in such areas as politics, economics, medicine and education. Many of these applications intend to find causal explanations that allow us to understand our environment.

However, although research methodology and decision making is supported by the study of association between the relevant variables, the existence of association does not necessarily imply a cause-effect relationship, but merely the existence of a co-variation between variables. Sometimes it is possible to find a high coefficient of correlation in variables in which there is no causal link (Ellet and Erickson, 1986).

Besides this epistemological difficulty, psychological research has shown that judging association is not an intuitive capability. Adults sometimes prefer to base their judgment on their previous beliefs about the type of association that ought to be between the variables that are going to be studied than on the empirical contingencies presented in the data.

Psychological research in this area has mainly concentrated on 2x2 contingency tables and provides us with valuable information concerning students' performance and strategies when judging association in this type of task. Nevertheless, from the point of view of Mathematics Education, the identification of students' preconceptions (Artigue, 1990; Confrey, 1990) is needed to plan adequate instruction. As Confrey (1990) pointed out, the relevance of the research on students' conceptions is because sometimes these conceptions differ in fundamental aspects from the scientific concepts that we try to teach and they are resistant to change in spite of instruction.

In this paper an experimental study with pre-university students is described, whose aim was to identify students' preconceptions concerning correlation in scatter plots. Students' judgments of association and strategies were used as empirical indicators of their conceptions. We present an original classification of the strategies, from a mathematical point of view, in which we have identified concepts and theorems in action as described by Vergnaud (1982). Correspondence analysis (Greenacre, 1984) is used to show the structure of students' strategies and the relationship between the strategies and the task variables.

PSYCHOLOGICAL RESEARCH ON STATISTICAL ASSOCIATION

Research on this topic has mainly concentrated in contingency tables, although there is some research concerning scatter plot, such as those from Erlick and Mills (1967) and Lane and cols. (1985).

A contingency table or cross-tabulation is used to present, in a summarized way, the frequencies in a population or sample, classified by two statistical variables. In its simplest form, when the variables only involve two different categories, it takes the format presented in Table 1.

Table 1. Typical format for the 2x2 contingency table

	B	Not B	Total
A	a	b	$a+b$
Not A	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

The study of reasoning about statistical association started with Inhelder and Piaget (1955), who considered the understanding of the idea of association as the last step in developing the idea of probability. So, understanding association has as prerequisites the concepts of proportionality, probability and the combinatorial capacity. Consequently, they only studied reasoning about association with children in their formal operation stage (IIIa and IIIb), proposing to the subjects the problem of the association between eyes and hair color, using a set of colored cards with drawings of faces as an experimental device (fair and brown hair; blue and black eyes).

With the aim of describing their results, Inhelder and Piaget used the scheme presented in Table 1, in which a, b, c and d represent the absolute frequencies in the four cells (fair hair, blue eyes; fair hair, black eyes; brown hair, blue eyes; brown hair, black eyes).

They found that at stage IIIa some adolescents only analyze the relationship between the favorable positive cases to the association (cell [a] in the Table 1) and the total number of data. In other cases they only compare the cell two by two. When they admit that the cases in call [d] (absence-absence) are also related to the existence of association, they do not understand that the cells [a] and [d] have the same meaning concerning the association, comparing [a] with [b] or [c] with [d] instead. This fact is explained because, although stage IIIa subjects can compute single probabilities, understanding association requires the consideration of the quantities $(a+d)$ as favorable to the association and $(b+c)$ as opposed to it and it is necessary to consider the relation:

$$R=\frac{(a+d)-(b+c)}{a+b+c+d}$$

between the difference of cases confirming the association $(a+d)$ and the other cases $(b+c)$ and all the possibilities. This is only produced at 15 years of age (stage IIIb), according to Piaget and Inhelder.

After Piaget and Inhelder, many psychologists have studied the adults' judgment of association in 2x2 contingency tables, using various kind of task and, consequently, it has been noted that subjects have a poor capacity for establishing a correct judgment about association. For example, Smedlund (1963) found that some adult students base their judgment only using cell [a] or by comparing [a] with [b].

The difficulty of this type of task is shown by the fact that, as Jenkins and Ward (1965) pointed out, even the strategy of comparing the diagonals in the table, considered as correct by Piaget and Inhelder, is only valid in tables having equal marginal frequencies for the independent variable. Nevertheless, in the research of Allan and Jenkins (1983) and Shaklee and Tucker (1980) this strategy was widely used by adults. For the general case, Jenkins and Ward have proposed, as the correct strategy, the comparison of the difference between the two conditional probabilities, $P(B|A)$ and $P(B|\bar{A})$:

$$\delta=\frac{a}{a+b}-\frac{c}{c+d}$$

Another point of interest in this topic is that Chapman and Chapman (1967) showed that there are common expectations and beliefs about the relationship between the variables that cause the impression of empirical contingencies. This phenomenon has been described as "illusory correlation," because people maintain their beliefs in spite of evidence of the independence of variables. Vital experience and cultural environment of the subject contribute to the formation of these theories that are used in interpreting data and facts around us. 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 the problem have on judging association.

The general conclusion is that when there is agreement between data and previous expectation people increase their confidence in the attribution of co-variation, but when data do not coincide with these theories there is a cognitive conflict and the accuracy in the perception of co-variation depends on the relative strength of the two sources of information.

As regards to scatter plot, Lane et al. (1985) argued that when a variable Y is linearly related to a variable X, then the relation between Y and X can be represented as $Y=bX+e$. In this equation b is the slope of the regression line and e is a random error whose mean value is equal to zero. The squared Pearson's correlation between X and Y (r^2) is determined from the following identity:

$$r^2=\frac{b^2S_x^2}{b^2S_x^2+S_e^2} \quad (1)$$

In this expression S_x^2 is the variance of X and S_e^2 the error variance; therefore Pearson's correlation is affected by the components of slope, error variance, and variance of X.

These three components of Pearson's correlation and the presentation format (graphical and tabular) were manipulated in three experiments by Lane et al. (1985). The first two experiments used naive subjects, and the third experiment used experts. The major conclusions were:

(1) scatter plots with the same value of Pearson's correlation were judged to posses different degrees of correlation when the correlations was besed on different combinations of the three components;

(2) when Pearson's correlation was held constant, the error variance was the main component affecting people's responses.

(3) graphical formats lead to higher performance than did tabular formats. This effect was larger for naive people than experts.

Erlick and Mills (1967) presented to their students a series of X and Y coordinates in sequence, asking them for an estimation of the correlation coefficient. Their results indicate a greater accuracy in estimating positive correlations, and positive bias of estimates when the sequence included a few large and many small deviations than when the data were homogeneous.

METHOD

The work we present here is part of wider research concerning the effect of a teaching experiment, based on problem solving and on the use of computers, on the learning of statistical association (Estepa, 1994; Estepa et al., 1994; Batanero et al., 1996).

Sample

The sample consisted of 213 students in the last year of secondary school (18 year old students). It is at this level where the topic of association is introduced in the Spanish curriculum. The questionnaire was given to the students before the instruction was started. So this study can be considered as a research on students preconceptions concerning statistical association (Artigue, 1990; Confrey, 1990). About half the students (113) were males and half females (100); 124 students had followed a scientifically oriented scondary curriculum and the rest a humanity's curriculum before they answered the questionnaire.

Questionnaire

The questionnaire included 3 items concerning scatter plots and is presented as an appendix. A pilot study of the questionnaire was performed with an additional sample of 51 students, which served to check the reliability and the coding system of the students' answers and to improve the final version. Since factor analysis of students' answer to the complete questionnaire had shown a multidimensional structure, a generalizability study (Brennan, 1982) was performed, instead of computing an index of internal consistency. For the complete questionnaire, composed by 10 items, including scatter plots, contingency tables and comparison of two samples, we obtained a generalizability index G=0.86 as a measure of possibility of extending our conclusions to the hypothetical item population and another index G=0.94 of generalizability to subjects population.

The following task variables were considered:

V1: Sign of the correlation: All three possible cases, direct, inverse and independence were used.

V2: Relationship between context and prior belief. The association suggested by the context of the problem and the empirical association presented in the table may coincide (theory agree with data) or not coincide (theory against data).

V3: Type of relationship between the variables: causal dapendence; interdependence or concordance.

V4: Value of the correlation coefficient. The precise values assigned in each of these variables in the different items are shown in Table 2.

Table 2. Values of task variables in the different items

Variables	Item 1	Item 2	Item 3
V1	Independence	Inverse	Direct
V2	Theory against data	Theory agree with data	No previous theory
V3	Interdependence	Indirect dependence	concordance
V4	0.11	-0.77	0.55

RESULTS AND DISCUSSION

Reading Scatter Plot

The complexity of a graphic is shown by the difficulty of reading and interpreting it. Curcio (1989) and Pichard, (1994) described reading a graph as the first level in graph comprehension, covering the reading of the axes, labels and the plot. In our research, we firstly studied the students' capacity to reading scatter plots.

To assess this capacity, we asked the students to read the value of ordinate when the abscissa was given as a data, in the first question of item 1, and to read the value of the abscissa when the ordinate was given as a data, in the second question of item 1.

Table 3. Frequency and percentage of answers in the reading the scatter plot

answer	reading the ordinate	reading the abscissa
Correct	184 (86.4)	181 (85.0)
Incorrect	25 (11.7)	26 (12.2)
No answer	4 (1.9)	6 (2.8)
Total	213 (100.0)	213 (100.0)

In table 3 we show the frequencies of the students' answers. 86.4 per cent of students were correct, in reading the ordinate and other 10 additional students provided the answer 10 and 12, very closed to the correct value.

85.0 per cent of students were correct, when reading the abscissa, and six additional students only gave one of the numbers 16 or 20 as the solution.

In short, about 90 per cent of students had reached the first level of graph comprehension in reading the scatter plot.

Analysis of Correlation Judgments

Once the data were collected, the arguments expressed by the students were categorized. Two dependent variables were considered in each item: the type of correlation perceived by the students (direct correlation, inverse correlation or independence) and the procedure employed by them to solve the proposed tasks. In table 4 we present the frequencies and percentages of the type of correlation perceived by the students.

In general the students were able to correctly judge the type of correlation between the variables. We can observe the greater difficult in Item 3. The spread in the scatter plot and the fact that the relationship between the variables was not due to a causal influence, but to an agreement between two classifications influenced the lower percentage of correct answers.

Table 4. Frequency and percent of type of correlation assessed by the students

Item	Independence	Direct	Inverse	No answer
1	*177 (83.1)	8 (3.8)	24 (11.3)	4 (1.9)
2	19 (8.9)	11 (5.2)	*182 (85.4)	1 (0.5)
3	137 (64.3)	*46 (21.6)	18 (8.5)	12 (5.6)

*correct answer

Classification of Students' Strategies in Scatter Plots from a Mathematical Point of View

In the previous analysis, we have shown how come students have correctly perceived the type of correlation in the data, which is an indicator of correct conceptions concerning correlation. But the idea of correlation, as is the case of many mathematical concepts, is not a simple idea. So, some students may have correct intuitive conceptions concerning some properties related to the idea of correlation mixed with some incorrect conceptions concerning different properties which have led them to choose an incorrect or partially correct strategy. All these conceptions need to be identified, to plan adequate instruction.

With the aim of identifying which of the properties linked to the concept of correlation were intuitively used correctly or incorrectly by our students, we analyzed the mathematical concepts and theorems in action that could be implied in the different procedures and the type of errors related to the same.

This task is, in the opinion of Vergnaud (1982; 1990) essential to the didactic analysis. However, Vergnaud recognizes that a theorem in action is not always a theorem, because then the student's recognition of the pertinence and truth of the theorem would be needed. But he also suggests that "*explicit concepts and theorems are only the visible part of the iceberg of conceptualization; without the hidden part, constituted by the operative invariants, the visible part would be nothing.*" (Vergnaud, 1990, p. 145). Taking this analysis as a base for classification, we have identified the following students' strategies, which will be described in the following paragraphs (a code will be included in brackets for each strategy, to be employed in the frequency table 5 and in the correspondence analysis of these strategies).

Correct strategies:

S1: **Global comparison** (GLOBAL): The student performed a correct global comparison of the relationships between the two variables. Students who employed this strategy could implicitly be using the following theorem T1: "*Independence means the invariance of the distribution of the dependent variable Y when conditioned with a value of X*". For example, in item 3 some students argued that both judges give similar punctuation to each student to justify dependence.

S2: **Increasing** (INCREASE): Using the increasing, decreasing or constant shape of the scatter plot to justify the type of dependence. Here, students implicitly could use the theorem T2: "*The increasing/decreasing shape of the scatter plot shows the existence of a direct/inverse correlation of variables.*"

Partially correct strategies:

S3: **Pattern comparison** (PATTERN): In item 3 the student compared the scatter plot with a linear function to correctly argument the association between the variables. In this strategy the theorem T3 could be implicit: "*When you can find a good fit of the scatter plot to a function, there is a moderate or high correlation between the two variables*". However, if the relationship is not linear, the coefficient may have low value.

S4: **Correct interpretation of isolated points** (CORRECTP): from which they deduced the correct type of correlation. For example, comparing the first and last point in item 2. However they do not use the complete information provided in the scatter plot showing a localist conception of correlation.

Incorrect strategies:

S5: **Incorrect interpretation of isolated points** (INCORRECTP). For example, comparing the drawn matches of the teams in first and in nineteenth place in item 1. Here was also shown the localist conception of the correlation.

S6: **Previous theories** (PTHEORY): The student based his judgments on previous beliefs about the relationshop between the variables, in spite of evidence of the data concerning a different type of correlation. Here we can identify the "illusory correlation" mechanism.

S7: **Other variables** (OTHERV): When, besides the independent variable, there are other variables that could effect the value of the dependent variable, as in item 1, some students considered there was no dependence between the two variables. These students interpreted the question concerning the dependence in a causal way. So, they show a causalist conception of correlation and required the elimination of rival explanations in order to admit dependence.

S8: **Uniformity** (UNIFORM): Students expected single-valued correspondence between the variables to consider the dependence, that is, they showed a deterministic conception of correlation.

S9: **Causality** (CAUSALTY): When in spite of empirical correlation the students argumented that there was no correlation because, as in item 3, the independent variable could not cause a direct influence on the dependent variable. We consider this answer as an indicator of a causalist conception of correlation.

S10: **Other wrong arguments** (OTHER): Other mistakes not included in the previous categories.

Table 5. Frequency and percent of strategies by implicit mathematical concept

Strategies	Item 1	Item 2	Item 3
GLOBAL	16(7.5)	3(1.4)	28(13.1)
INCREASE	15(7.0)	134(62.9)	8(3.8)
PATTERN	0	0	24(11.3)
CORRECTP	49(23.0)	27(12.7)	4(1.9)
INCORRECTP	4(1.9)	3(1.4)	25(11.7)
PTHEORY	22(10.3)	15(7.0)	6(2.8)
OTHERV	68(31.9)	1(0.5)	0
UNIFORM	14(6.6)	1(0.5)	21(9.9)
CAUSALITY	0	0	50(23.5)
OTHER	25(11.7)	29(13.6)	47(22.1)

In table 5 we present the percentages of the different strategies in each item. We can observe many correct strategies in Item 2, in which the correlation is very high. Concerning the partial correct strategies, the higher frequency was the correct judgment based in isolated points. Nibett and Ross (1980) suggested that the information related to consensus is needed to obtain a correct judgment of the association between two variables. This information indicates to what extent a particular individual is an exception in a group and when its behavior is shared by most cases in the group. In spite of this Kahneman and Tversky (1973) showed that, typically, subjects do not consider the importance of the bases rated of a particular behavior in a population or in a group. On the contrary, they ignore the consensus information. This fact could explain the strategies S4 and S5 in our students, who based their judgments in isolated points.

In the incorrect strategies the higher frequency corresponded to strategies S6, S7 and S9, in which there was an underlying causal conception of correlation. According Tversky and Kahneman (1982), in the judgment of association and causality, the conditional probabilities P(X/D) of an event X, give some data D are investigated. From the psychological point of view, the subject may perceive different type of relationship concerning this conditional probability.

When D is considered as the cause for the occurrence of X, we call D a causal data. If X is perceived as the cause of D, D is a diagnostic data. In the case that D and X were considered as consequences of other different factors, D is an indicative data of X. Finally, when we believe that there is no relationship between D and X, we call D an incidental data. In the normative study of conditional probability and of association is not needed to distinguish among these types of relationships. Neverthe-less, these differences may be important from a psychological point of view.

In our opinion, some of our students did not attribute a symmetrical role to both variables. Instead, they used treir expectations concerning the type of relationship that ought to be between the variables, showing a causalist conception of correlation.

Correspondence analysis between strategies and items

According to Vergnaud (1990), part of the meaning of mathematical concepts depends on the situations in which these concepts are applied. So, we can expect a shift in the subject strategies in our study, depending on the task to be solved. As the main task variables in our research have a mathematical nature, we must expect a dependence of the students' strategies on these variables.

To describe this dependence, a correspondence analysis (Greenacre, 1984) of Table 5 was performed, using BMDP package. Task variables of the items were used as supplementary variables, with the aim of improving the interpretation.

The Chi-square statistic value was highly singnificant and result of the analysis showed a bidimensional structure. Representational quality of rows and columns was greater than 0.707, which is very high; this fact validates the categorization chosen to describe the structure of the strategies that the students in our sample have used to solve the proposed problems. In the following paragraphs we will describe the result of the analysis of the two factors that were identified. For each row and columns of Table 5, we will denote by x the value of the coordinate of this row on the factor and by r the value of the correlation coefficient between the row and the factor.

The first factor (54.3% of the total inertia) showed the differentiation of student's strategies in Item 3 (x = 1.008; r = 0.958) with respect to the other two items that had negative coordinates on the axis. In this items the most frequent strategies were the comparison with a pattern (PATTERN; x = 1.368; r = 0.942); the causal argumentations (CAUSAL; x = 1.368; r = 0.952), employing isolated bad interpreted points (INCOR-RECTP; x = 0.927, r = 0.960) and considering a uniform graph (UNIFORM; x= 0.717, r = 0.806).

On the contrary, in lems 1 and 2 students have preferred the correct interpretation of isolated points (CORRECT; x = -0.506, r = 0.537); looking to the increasing or decreasing of the scatter plot (INCREASE; x = -0.506, r = 0.537 and previous theories (x = -0.355, r = 0.598). For us this is an evidence of the causalist conception of association, because the students have shifted their strategy for Item 3, in which the correlation was not due to causal influence, but to concordance.

The second factor (42% of inertia) opposes independence (Item 1; x 0.852; r 0.895) to the other two items, with negatives coordinates on this axis. The strategies of considering that the existence of other variables that could influence the dependent variable proved not association (OTHERV; x = 1.323; r = 0.915) and correct interpretation of isolated points (CORRECTP; x = 0.470; r = 0.463) were linked to the independence. On the contrary, there was a greater use of strategy of studying the increase of the graph (INCREASE; x = -0.756; r = 0.476) in case of association, because this is a property of the scatter plot easily grasped by the students in case of high or moderate correlation.

CONCLUSIONS

In the previous sections we have presented an experimental study of students' strategies in judging correlation in scatter plots. The classification of these strategies, from a mathematical point of view, has allowed us to identify intuitively correct and partially correct strategies that are indicators of correct or partially correct conceptions concerning statistical association, in particular:

-using the increasing, decreasing or constant shape of the scatter plot to justify the type of dependence;

-comparing the scatter plot with a known function - for example lineal - to correctly argument the association between the variables.

In the same way, incorrect strategies and judgments have served to identify the three following incorrect conceptions:

-**Determinist conception of correlation:** some students did not admit exceptions to the existence of a relationship between the variables. The expected a correspondence that assigns only a value in the dependent variable for each value of the independent variable. When this is not so, they considered there was not dependence between the variables.

-**Localist conception of correlation:** Students often based their judgment in only part of the data provided in the scatter plot. If this partial information served to confirm a given type of correlation, they adopted this type of association in their answer.

-**Causalistic conception of correlation:** when the students identified correlation and causality.

The correspondence analysis performed to relate strategies and task variables of the items have shown two different factors. The first factor points to the confusion between correlation and causality that has provoked many incorrect strategies in the concordance case, as a means to justify the students' previous beliefs aboaut the correlation suggested by the context of the item. The second factor differentiates the strategies used for justifying independence from those used to admit correlation.

All these findings show the complexity of a topic simple in appearance. Because of this fact and the relevance of understanding the idea of association, we deduce the need to reinforce the teaching of this topic in university level and in the last years of secondary education. This will require informing teachers of the mathematical, epistemological, psychological and educational aspects of the topic, including information about students' preconceptions as enunciate in this paper. Also needed is an adequate design of didactical situations in which students need to make explicit their misconceptions, with the aim of provoking in them a cognitive conflict to overcome these misconceptions.

Acknowledgment: This report forms part of the Project PS91-0114.1, granted by Dirección General de Investigacion Cientifica y Técnica (M. E. C.). Madrid. And Group Jaén of Didactic, code 1349, Junta de Andalucía.

REFERENCES

Allan, L. G. & Jenkins, H. M., (1983), The effect of representations of binary variables on judgment of influence. *Learning and motivation, 14*, 381-405.

Alloy, L. B. & Tabachnik, N.,(1984), Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational imformation. *Psychological review, 91*, 1112-149.

Artigue, M.,(1990), Epistemologie et Didactique. *Recherches en Didactique des Mathématiques, 10(2-3)*, 241-286.

Batanero, C., Estepa, A., Godino, J. D. Y Green, D. R., (1996), Judgments of association in contingency tables. *Journal for Research in Mathematics Education, 27(2)*, 151-169.

Brennan, R. L.,(1982), *Elements of Generalizability Theory*. ACT Publications, Iowa.

Chapman, L. J. & Chapman, J. P.,(1967), 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 in students conceptions in mathematics, science and programming. *Review of Research in Education, 16*, 3-35.

Curcio, F. R.,(1989), *Developing Graph Comprehension*. Reston: (Virginia): Mational Council of Teachers of Mathematucs.

Ellet, F. S. & Erickson, D. P.,(1986), Correlation, partial correlation and causation. *Synthese, 67*, 157-173.

Erlick, E. & Mills, R. G.,(1967), Perceptual quantification of conditional dependency. *Journal of Expermental Psychology. 73(1)*, 9-14

Estepa, A.,(1994), *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores.* (Preconceptions on statistical association and their evolution as result of a computer based teaching). Ph. D. University of Granada (Spain).

Estepa, A. & et al.(1994), Judgments of association in contingency tables: an empirical study of students' strategies and preconceptions. In F. Matos and J. F. Ponte (Eds.). (1994). *Proceeding of Eighteenth Annual Conference of the International Group for the Psychology of Mathematics Education (PME 18). Vol 2*, 312-320. Lisbon: Program Committee.

Greenacre, M. J., (1984), *Theory and applications of correspondence analysis*. Academic Press, London.

Inhelder, B. & Piaget, J.,(1955), *De la logique de l'enfant á la logique de l'adolescent* (From the logic of the chid to the logic oc the adolescent). Presses Universitaries de France, Paris.

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

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

Kahneman, D. & Tversky, A.,(1973), On the psychology of prediction. In D. Kahneman, P. Slovic & A. Tversky (Ed.), *Judgment under uncertainty: Heuristics and biases*, New York: Cambridge University Press. 48-68.

Lane D. M., Anderson, C. A. & Kellan, K. L.,(1985), Judging the relatedness of variables: The psychophisics of covariation detection. *Journal of Experimental Psychology. Perception and Performance, 11, 5*, 640-649.

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

Pichard, J. F.(1994). Approche epistemologique de la complexite des graphiques statistiques. *Proceeding of Fourth International Conference on Teaching Statistics*, 486. Rabat (MOROCCO): The National Institute of Statistics and Applied Economics.

Shaklee, H. & Tucker, D.,(1980), A rule analysis judgments of covariation between events. *Memory and Cognition, 8*, 459-467.

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

Tversky, A. & Kahneman, D.,(1982), Causal schemes in judgments under uncertainty. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge Universitary Press. 117-128.

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.

Vergnaud, G.,(1990), La théorie des champs conceptuells. *Recherches en Didactiques des Mathématiques, Vol. 10(2-3)*, 133-170.

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

Departamento de Didáctica de la Ciencias
Facultad de Humanidades Y Ciencias de la Educación.
C/Virgen de la Cabeza, n. 4
Universidad de Jaén
23008 Jaén
Spain

APPENDIX: QUESTIONNAIRE

ITEM 1. In the attached scatter plot we have represented the place occupied by "primera división" team of Spanish football league (years, 1987/1988) and the matches drawn. A football team obtain two points for each match won, one point for each match drawn, and zero point for each match lost. The final classification of football league is made according to the total number of points obtained for each team.

1. How many matches were drawn by the team classified in the tenth place?
2. What place obtained the teams with five drawn matches?
3. Do you believe that the relationship between the place occupied for every team and the number of drawn matches is direct, inverse or that there is no relationship? (Explain you answer).

ITEM 2. In a sociological study, data relative to daily consumption of animal protein and birth rate of different countries were collected. There data were represented in the attached scatter plot.

Do you think that the relationship between daily consumption of animal protein and birth rate of there different countries is direct, inverse or that there is no relationship at all? (Explain your answer).

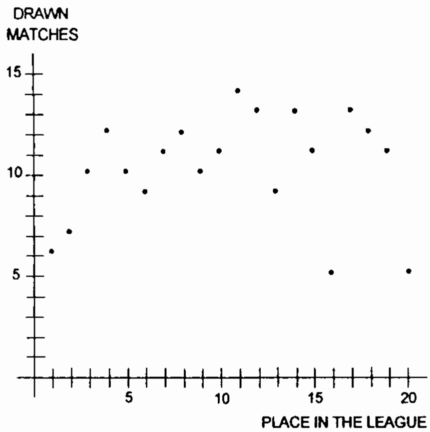

Figure 1

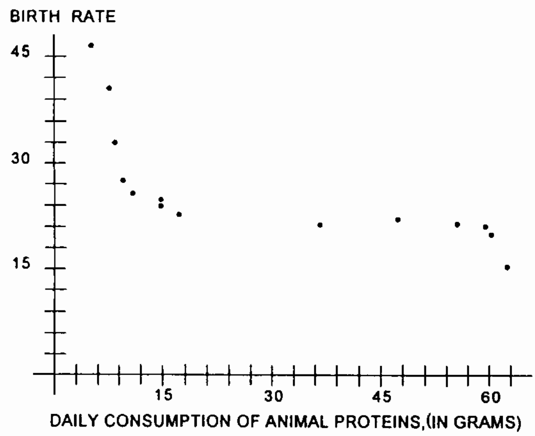

Figure 2

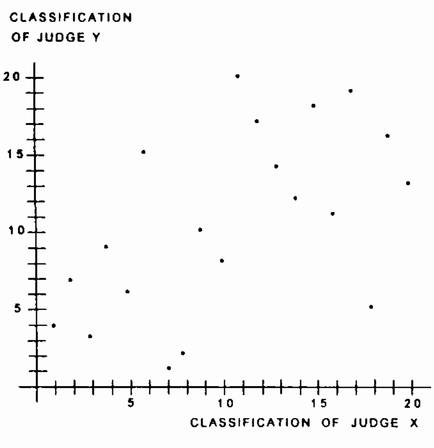

Figure 3

ITEM 3. Two judges (judge X, judge Y) qualified twenty students who took part in a project competition in accordance with their opinion. We have represented in this plot the place that each student was given in both judges' classification. We would like to study if there is some relationship between the place assigned to the same participant by the two judges (if the judgment of both judges is or not related). Observe the diagram and answer:

Do you believe that the relationship between the place assinged to the same student by both judges is direct, inverse or that there is no relationship at all? (Explain your answer).