



Judging the Relatedness of Variables: The Psychophysics of Covariation Detection

David M. Lane, Craig A. Anderson, and Kathryn L. Kellam
Rice University

Previous research on how people judge the relation between continuous variables has indicated that judgments of scatterplots are curvilinearly related to Pearson's correlation coefficient. In this article, we argue that because Pearson's correlation is composed of three distinct components (slope, error variance, and variance of X) it is better to look at judgments as a function of these components rather than as a function of Pearson's correlation. These three components of Pearson's correlation and presentation format (graphical and tabular) were manipulated factorially in three experiments. The first two experiments used naive subjects, and the third experiment used expert subjects. The major conclusions were (a) scatterplots with the same value of Pearson's correlation are judged to possess different degrees of relation if the correlations are based on different combinations of the three components; (b) with Pearson's correlation held constant, the error variance is the most important component; and (c) graphical formats lead to higher judgments of relatedness than do tabular formats, with this effect being larger for naive than for expert observers. It was also concluded that attempts to determine the psychophysical function between Pearson's correlation and judgments of relatedness are of questionable value.

Although judgments of covariation between dichotomous variables have been studied extensively (see Arkes & Harkness, 1983, for a review), much less attention has been directed toward judgments of covariation between continuous variables. Central to understanding how judgments of the relation between variables are made is the determination of the psychophysical function between measurable characteristics of the relation and human judgments of it. Two studies have addressed this question by seeking to determine the psychophysical function between Pearson's correlations and subjects' judgments of covariation (Cleveland, Diaconis, & McGill, 1982; Jennings, Amabile, & Ross, 1982), and both found judgments to be a positively accelerated function of Pearson's correlation. Jennings et al.'s data were fit well by the function $1 - (1 - r^2)^{1/2}$, and, although this function did not fit the data obtained by Cleveland et al. quite as well,

the fit was still reasonably good. Thus, although the precise mathematical form of the psychophysical function may be in doubt, the general shape of the function appears to be clear.

Despite this initial success in finding a function of Pearson's correlation that fits human judgments reasonably well, it is probable that some aspects of the judgment process are missed when the relation between the variables is summarized by Pearson's correlation. Specifically, because Pearson's correlation is itself based on the three components of a linear relation between variables (slope, the variance of X and error variance as explained below), some information about the individual effects of these components may not be contained in the function relating values of Pearson's correlation to human judgments. The problem of analyzing judgments of covariation in terms of these three components is similar to the problem of analyzing judgments of coldness in terms of the two components of heat loss (air temperature and wind velocity) in that in both cases there are underlying components and a commonly accepted method for combining them: Pearson's correlation in the former case and the wind-chill factor in the latter. Naturally, if one were interested in judgments

We thank Alma Anorga for her help in conducting Experiment 1.

Requests for reprints should be sent to David M. Lane at the Department of Psychology, P.O. Box 1892, Rice University, Houston, Texas 77251.

of coldness, one would look at these judgments as a joint function of air temperature and wind velocity and not only as a function of the wind-chill factor.

If a variable Y is linearly related to a variable X , then the relation between Y and X can be represented as $Y = bX + e$, where b is the slope and e is a random variable with a mean of zero. The variance of Y , as well as the Pearson's correlation between X and $Y(r)$, is determined from the slope (b), the variance of $X(S_x^2)$ and the error variance (S_e^2): The variance of Y is $b^2S_x^2 + S_e^2$; Pearson's correlation squared is the variance explained by $X(b^2S_x^2)$ divided by the variance of Y . This latter expression, which is shown as Equation 1, illustrates how Pearson's correlation is made up of the components of slope, error variance, and variance of X :

$$r^2 = b^2S_x^2 / (b^2S_x^2 + S_e^2). \quad (1)$$

In order for judgments of covariation to be a monotonic function of Pearson's correlation, these judgments must be the same for all data having the same value of Pearson's correlation regardless of the values of the individual components. It is not necessarily the case, however, that people's judgments of scatterplots with the same Pearson's correlation would be the same. For example, values of slope, error variance, and variance of X of 2, 1,000, and 400 lead to a Pearson's correlation of .78 as do values of 4, 4,000, and 100. It is an empirical question as to whether judgments of these two relations would be equal. If they are not, then one could find a set of two combinations of the three components for which the combination that results in higher judgments has the lower Pearson's correlation simply by changing one of the three components of the set with the higher judged relation very slightly in the direction that would lower Pearson's correlation.

Although, to our knowledge, no one has manipulated the components of Pearson's correlation in a covariation assessment paradigm, research on single-cue and multiple-cue probability learning has been concerned with the components of slope and error variance.¹ For example, Brehmer and Lindberg (1970) found that increasing the error variance while holding slope and variance of X constant had a detrimental effect on subjects' learning. Although this study was subsequently criticized by Brehmer (1973) for having confounded er-

ror variance with Pearson's correlation, we do not believe this criticism is valid. Just as it would not make sense to criticize a study finding that judgments of cold decrease as temperature decreases for confounding temperature and wind chill, it does not make sense to criticize Brehmer and Lindberg's study for confounding error variance and Pearson's correlation. Pearson's correlation is not a variable of the same class as error variance, but rather a conglomeration of error variance, slope, and variance of X .

Two studies have compared people's ability to learn relations having the same value of Pearson's correlation but differing in slope and error variance. Uhl (1966) found that changing either error variance or slope did not affect learning rate as long as there was a change in the other variable sufficient to maintain Pearson's correlation at a constant level. Contrary results were obtained in a study by Eade (1967) in which a combination of high error variance and high slope led to poorer performance than did a combination of low error variance and low slope even though Pearson's correlation was the same in both instances. One possible reason for the discrepancy between the two studies is that Eade used a wider range of slopes and error variances and found the effect only for the most extreme combinations. Thus, there is evidence that at least in some situations, performance cannot be predicted from the value of Pearson's correlation as well as it can be predicted from the underlying components of Pearson's correlation.

The major purpose of the present research was to investigate the effect of the three components of Pearson's correlation on subjects' judgments of covariation. In order to compare the way in which these components affect subjects' judgments with the way they are combined in the computation of Pearson's corre-

¹ In single-cue probability learning, subjects are given a series of trials in which they are presented with a value of X and asked to predict the value of Y . After responding, they are shown the actual value of Y . Stimuli are constructed so that the values of Y are not perfectly predictable from X . Most of the research done with this paradigm has been concerned with the effect of the degree and type of relation between X and Y on learning (see Naylor & Domine, 1981, for a review). Multiple-cue probability learning tasks differ from single-cue tasks only in that Y is to be predicted from several variables rather than just one.

lation, we included more than one combination of these components for a given value of Pearson's correlation.

We were also interested in the effect of the format of presentation (graphical vs. tabular) and how it might interact with the various components of relatedness. Although both graphical formats (Cleveland et al., 1982; Erlick & Mills, 1967) and tabular formats (Jennings et al., 1982) have been employed in studies of judgments of relatedness, they have never been compared directly.

Experiment 1

Method

Subjects. Thirty-nine undergraduate psychology students participated in this study for course credit.

Materials. The stimulus materials were packets of either 40 tables of data for X and Y variables or 40 graphs showing scatterplots of the same data. Each table contained nine X - Y number pairs presented in ascending order on the X variate; each graph displayed the same nine data points as the corresponding table. No numerical values were presented on the graphs; the X and Y axes were unlabeled and were simply anchored *low* and *high*. Below each table or graph was a point rating scale with ticks at 10-point intervals and numeric anchors at 0 and 100.

The data portrayed in the various tables or graphs differed on three variables: slope, with a value of two or four; error variance, with a value of 1,000 or 4,000; and the variance of X , with a value of 100 or 400. These six values resulted in the eight possible combinations shown in Table 1. Correlations of .53 and .78 were each produced by three combinations of the variables. One combination produced a correlation of .30, and one produced a correlation of .93. Five different sets of these eight combinations were used. Thus, there was a total of 40 stimulus sets. Note that the differences between the levels of these factors are in some sense equivalent because changes in the level of any one factor produces equivalent changes in r .

An example of the tabular and graphical modes of presentation is shown in Figure 1. The slope is 4, the error variance is 4,000, and the variance of X is 400.

The nine values of X for each data set were generated as follows. First, the fifth highest X was set equal to 0. Then, four displacements from 0 were generated by adding a random component (rectangularly distributed from $-.1$ to $.1$) to the expected normal score for the serial position in question. These displacements constituted the four highest values of X ; the arithmetic inverses of these displacements constituted the lowest four numbers. Therefore, the X values were symmetrically distributed. Finally, 50 was added to each value of X , and these values were rescaled so that they had the desired variance (100 or 400, depending on the condition). The 10 values of Y were computed by multiplying each value of $(X - 50)$ by the desired slope (2 or 4), adding an error component, and then adding 50. The error component for the lowest value of X was the same as that for the highest value; the value for the second highest was the same as that for the second lowest, and so

Table 1
Correlations Contained in the Stimulus Data Sets

Variance of X	Slope			
	Low (2)		High (4)	
	LEV (1,000)	HEV (4,000)	LEV (1,000)	HEV (4,000)
Low (100)	.53	.30	.78	.53
High (400)	.78	.53	.93	.78

Note. LEV = low error variance; HEV = high error variance.

forth. This pattern was necessary in order to keep the error components from changing the slope. The error components had a mean of 0 and were equal to the values of the expected normal scores for five numbers scaled to have the desired variance (1,000 or 4,000).

Procedure. Subjects were randomly assigned to either the graphical or tabular format condition and were tested individually or in small groups of from two to six. The same randomly determined order of stimulus presentation was used for all subjects. Packets were distributed, and subjects read their instructions from the packet's cover sheet as the experimenter read these same instructions out loud:

This is an experiment to find out how people such as you detect and estimate the relationship between two variables, X and Y , in a table or in a scatterplot. You will measure relationships on a scale of 0 to 100. Zero means no relationship and 100 means a perfect linear relationship. Rate each table or scatterplot according to your subjective assessment of the relationship. Please mark an "X" directly on the scale to indicate your assessment. There is no right answer. If you have any questions, feel free to ask the experimenter. Please work carefully and rapidly, taking no more than 1 minute per problem.

Any questions from subjects about how one estimates a relation or what constitutes a relation were answered in vague terms by the experimenter to avoid emphasizing any of the three components to the subjects.

Results

The mean judgments of relatedness as a function of slope, error variance, and the variance of X are presented in Table 2. The main effect of each of these factors reveals whether or not it affected subjects' estimates.

An unweighted means analysis of variance was performed. Significant main effects of error variance and variance of X were obtained, $F_s(1, 37) = 31.98$ and 16.17 , $p_s < .001$. Subjects judged the relations to be stronger when error variance was low ($M = 22.1$) than when it was high ($M = 16.9$). Subjects also judged

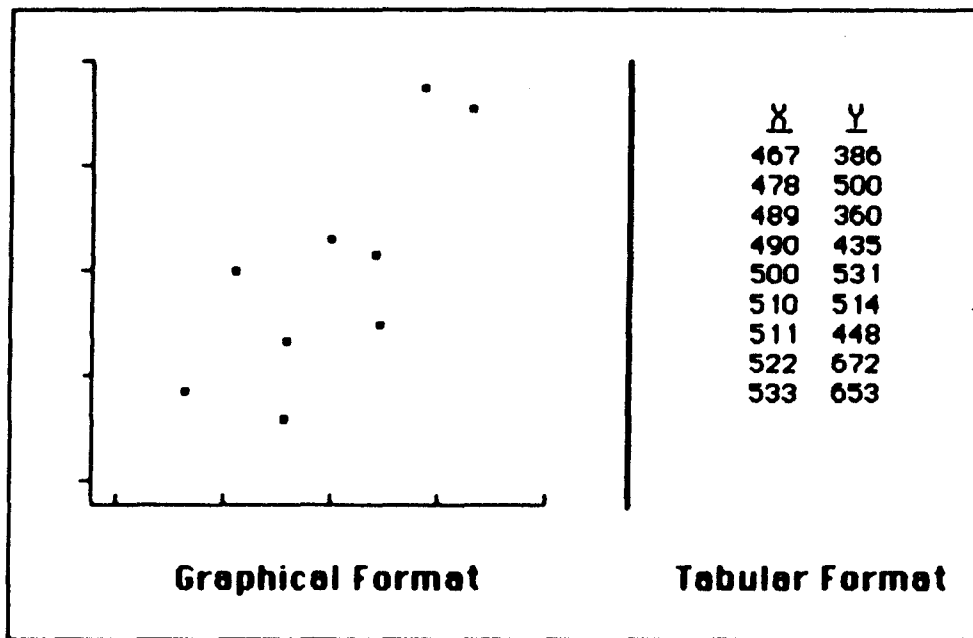


Figure 1. An example of the graphical and tabular methods of presentation. (The slope is 4; the error variance is 4,000; the variance of X is 400; and r is .78.)

relations to be stronger when variance of X was high ($M = 21.0$) than when variance of X was low ($M = 17.9$). In addition, subjects were much more sensitive to differences in variance of X when error variance was low than when error variance was high, $F(1, 37) = 7.51, p = .009$. Interestingly, there was no main effect of slope, $F(1, 37) = 1.12, p = .297$.

Overall, subjects judged the degree of relatedness to be higher in the graphical format than in the tabular format, $M_s = 24.3$ and 14.7 respectively, $F(1, 37) = 23.87, p < .001$. In general, the effects of the three variables

were stronger in the graphical than in the tabular format. There was a significant Format \times Variance of X interaction, $F(1, 37) = 15.77, p < .001$ and some hint of a Format \times Error variance interaction, $F(1, 37) = 3.85, p = .057$. Finally, there was a significant Format \times Slope \times Variance of X interaction reflecting the fact that the Slope \times Variance of X interaction was bigger in the graphical format than in the tabular format condition, $F(1, 37) = 5.97, p = .019$. No other effects approached significance.

The major purpose of examining subjects' sensitivity to slope, error variance, and variance of X was to see if people are affected by these components in the same way as is Pearson's correlation coefficient. Mean judgments as a function of Pearson's correlation are shown in Figure 2. The three plus and/or minus signs associated with each point indicate the levels of slope, error variance, and variance of X , respectively, with plus signs indicating that the value is associated with a higher correlation and minus signs indicating that the value is associated with a lower correlation. Therefore, high slope, low error variance, and high variance of X is represented as (+++); high slope, low error variance, and low variance of X is represented as (++-); high slope, high error variance, and high variance of X is represented as (+-+) and so forth. It can be

Table 2
Mean Judgments of the Relation Between X and Y , Experiment 1

Data type and variance of X	Slope			
	Low		High	
	LEV	HEV	LEV	HEV
Graphical				
Low	22.2	17.0	25.2	20.3
High	32.4	23.6	31.3	22.2
Tabular				
Low	14.4	14.0	16.6	13.6
High	16.0	11.8	18.4	12.6

Note. LEV = low error variance; HEV = high error variance.



Figure 2. Mean judgments of the relation between X and Y as a function of presentation format (Graphical [G] or Tabular [T]) and slope, error variance, and variance of X of the stimulus data, Experiment 1. (Note: The three plus and/or minus signs associated with each point represent the levels of slope, error, and variance, respectively. Plus signs indicate that the value is associated with a higher correlation [high slope, low error, or high variance], and minus signs indicate the reverse [low slope, high error, or low variance].)

seen in Figure 2 that judgments of scatterplots possessing the same Pearson's correlations were judged to have different degrees of relatedness.

For the graphical format condition, judgments of relatedness in the low slope, low error variance, high variance of X condition were quite a bit higher than judgments in either the high slope, low error variance, low variance of X condition or the high slope, high error variance, high variance of X condition even though Pearson's correlation was .78 in all three cases. Using the Newman-Keuls criterion in comparing the means of these three conditions (a separate error term was computed for each comparison), significant differences ($p < .01$) were found between the first and last of these conditions and between the first and the second of these conditions. The difference between the second and the third conditions was not significant ($p = .18$).

The pattern obtained with the tabular format was similar to that found with the graphical format. Judgments in the high slope, low error variance, low variance of X condition were significantly higher than in the high slope,

high error variance, high variance of X condition ($p < .01$). The difference between the low slope, low error variance, high variance of X condition and the high slope, high error variance, high variance of X condition did not quite reach conventional levels of significance ($p = .052$).

Discussion

Two clear-cut conclusions can be drawn from the present data. First, people judge scatterplots portraying the same degree of relation as defined by Pearson's correlation differently if the component parts of the correlation are different. It appears that the lower the error variance, the higher the judgment of the relation even if the slope and/or the variance of X are reduced so as to maintain the value of Pearson's correlation at a constant level. In this sense, people are influenced more by error variance than by either slope or variance of X relative to how these factors influence Pearson's correlation, a result consistent with those of Eade (1967). We do not consider these effects to be biases because in no sense does

Pearson's correlation represent the "true" measure of relation. Second, people perceive variables to be more highly related if the data are portrayed in a graphical format than if they are portrayed in a tabular format.

One intriguing finding was the lack of a significant main effect of slope. Because this is a rather counterintuitive finding and because of the hazards inherent in accepting the null hypothesis, we decided to attempt to replicate it before trying to interpret it.

Experiment 2

Method

Forty undergraduate psychology students participated in this study. Subjects were randomly assigned to either the graphical or the tabular format conditions, with the constraint that an equal number of subjects was assigned to each condition. The stimulus materials and procedure were exactly the same as in Experiment 1.

Results

Table 3 presents the mean judgments for all conditions. There were sizable main effects of all three factors. Subjects' judgments were higher when the slope was high ($M = 40.4$) than when the slope was low ($M = 34.8$), $F(1, 38) = 17.74, p < .001$. Relations with low error were judged higher than relations with high error, $M_s = 44.0$ and 31.2 , respectively, $F(1, 38) = 62.59, p < .001$. Finally, subjects judged relations with high variance of X as stronger than those with low variance of X , $M_s = 41.8$ and 33.4 , respectively, $F(1, 38) = 33.66, p < .001$.

The Error \times Variance of X interaction found in Experiment 1 was replicated in Experiment 2, $F(1, 38) = 10.60, p = .002$, as was the main effect of format, $F(1, 38) = 5.96, p = .019$. A number of other less interesting interactions were also significant: Format \times Error, $F(1, 38) = 28.98, p < .001$; Format \times Variance of X , $F(1, 38) = 22.77, p < .001$; Slope \times Variance of X , $F(1, 38) = 25.30, p < .01$; and Format \times Slope \times Variance of X , $F(1, 38) = 6.81, p = .012$. The Format \times Slope interaction reflects the fact that subjects were more sensitive to slope in the graphical format condition. The Format \times Slope \times Variance of X interaction reflects the fact that the Slope \times Variance of X interaction occurred only in the graphical format condition. The patterns and interpre-

Table 3
Mean Judgments of the Relation Between X and Y , Experiment 2

Data type and variance of X	Slope			
	Low		High	
	LEV	HEV	LEV	HEV
Graphical				
Low	35.0	20.6	55.2	35.9
High	60.9	36.2	69.3	41.8
Tabular				
Low	31.3	30.4	30.5	28.1
High	35.0	28.8	34.6	27.8

Note. LEV = low error variance; HEV = high error variance.

tations of the other interactions were much as they were in Experiment 1. Other effects were not significant.

More important, as in Experiment 1, different combinations of components producing the same Pearson's correlation were judged to portray different degrees of relatedness. As can be seen in Figure 3, this aspect of the results of Experiment 1 was almost perfectly replicated. The one outstanding difference was that the absolute level of judged relatedness was quite a bit higher in Experiment 2 than it was in Experiment 1.

Significant differences were found in the graphical format conditions between the judgments from the three combinations that resulted in a Pearson's correlation of .78. Newman-Keuls tests revealed that judgments in the low slope, low error variance, high variance of X condition were significantly higher ($p < .01$) than the judgments in the high slope, high error variance, high variance of X condition and that judgments in the high slope, low error variance, low variance of X condition were significantly higher than those in the high slope, high error variance, and high variance of X condition. The former comparison indicates that error variance was relatively more important than slope, whereas the latter comparison indicates that error variance was relatively more important than variance of X . The difference between the low slope, low error variance, high variance of X condition and the high slope, low error variance, low variance of X condition was not significant ($p = .18$).

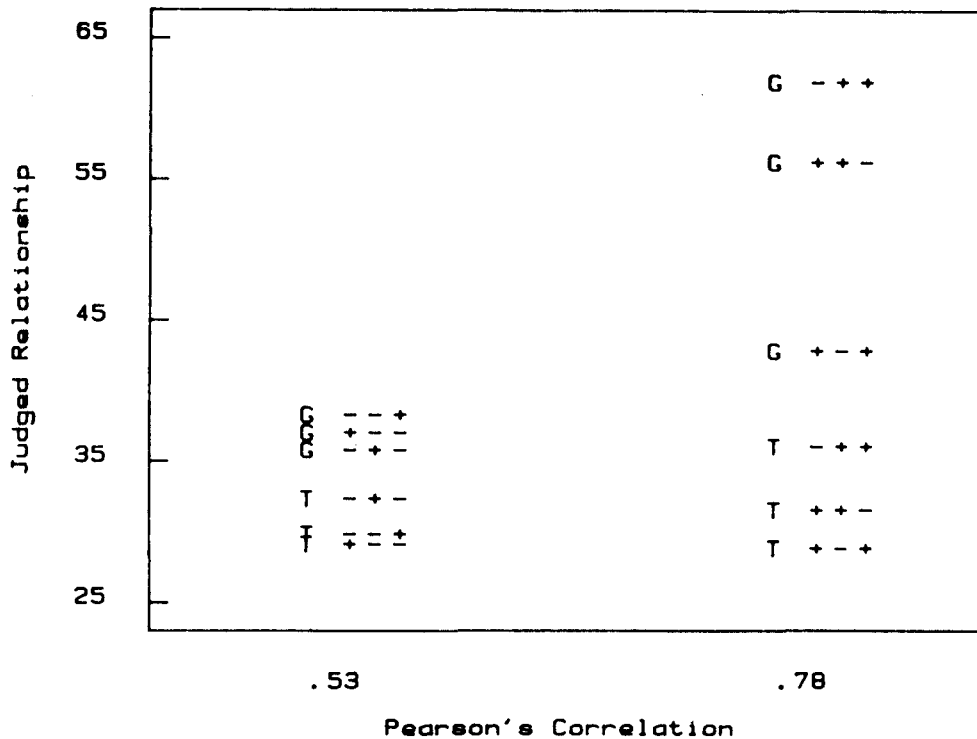


Figure 3. Mean judgments of the relation between X and Y as a function of presentation format (Graphical [G] or Tabular [T]) and slope, error, and variance of X of the stimulus data, Experiment 2.

Discussion

Slope had a significant effect on subjects' judgments in Experiment 2, contradicting the null result in Experiment 1. However, it should be noted that the effect of slope was in the same (and expected) direction in both experiments; it simply failed to reach significance in Experiment 1. Otherwise, the results of Experiment 2 were very similar to those of Experiment 1. In both experiments subjects' judgments were higher when the relations had low error variance even though the Pearson's r was unchanged due to changes in slope or variance of X . Also, graphical presentation produced larger estimates in general as well as stronger effects of the three component factors than did tabular presentation.

These data, therefore, provide a fairly clear picture of how statistically naive subjects perform on covariation detection tasks. Briefly, people are affected by all three components of linear relations, but are not influenced by them in precisely the same way that the Pearson's r is. Would statistically sophisticated subjects show the same pattern as naive subjects, or would their judgments reflect a Pearsonian combination? In Experiment 3 we examine this issue.

Experiment 3

Method

Subjects. Twenty-five PhD or EdD professionals in psychology, economics, and education volunteered to take part in this study as part of a statistically expert sample. Subjects had taken an average of 5.8 statistics courses in college, with 2.75 of these courses dealing with correlations (medians of 4 and 2.5, respectively). They considered themselves to have a better grasp of statistics than 91% of the general population and 61% of their professional peers.

Procedure. The stimulus materials were exactly the same as in Experiments 1 and 2. Subjects were contacted in person or by phone and asked to participate. The materials, with cover-page instructions, were then delivered to those who consented. Subjects completed the materials in private and returned them to the experimenter's departmental mailbox or to a neutral third party. As in the previous experiments, subjects were instructed to give their subjective assessments of the relations.

Results

Table 4 presents the mean judgments of relatedness. Once again, an unweighted means analysis of variance was performed, and there were large main effects of the three components of correlations. The high and low slope means were 47.9, and 34.3, $F(1, 23) = 52.88$, $p < .001$. The high and low error variance means were 30.3 and 51.9, $F(1, 23) = 158.58$,

Table 4
Mean Judgments of the Relation Between
X and *Y*, Experiment 3

Data type and variance of <i>X</i>	Slope			
	Low		High	
	LEV	HEV	LEV	HEV
Graphical				
Low	34.8	10.7	55.4	27.1
High	58.9	32.7	72.5	48.1
Tabular				
Low	37.1	24.2	44.1	28.2
High	48.7	27.5	63.8	43.8

Note. LEV = low error variance; HEV = high error variance.

$p < .001$. The high and low variance of *X* means were 49.5 and 32.7, $F(1, 23) = 73.59$, $p < .001$. However, there was no main effect of format ($F < 1$). The following interactions were all significant: Format \times Error variance, $F(1, 23) = 5.79$, $p < .05$; Format \times Variance of *X*, $F(1, 23) = 4.72$, $p < .040$; and Format \times Slope \times Variance of *X*, $F(1, 23) = 15.31$, $p < .001$. The first two of these interactions followed the same patterns as in Experiments 1 and 2, and reflected the fact that both error

variance and variance of *X* had larger effects with the graphical presentation. The form of this last interaction was also the same as in the first two experiments. Basically, an increase in slope had a smaller effect in the high variance of *X* condition than in low variance of *X* condition for the graphical format, whereas the opposite was true for the tabular format. No other effects were significant.

As can be seen in Figure 4, the major result of Experiments 1 and 2 was replicated: Different combinations of components producing the same Pearson's correlation were judged to have different degrees of relatedness. Moreover, several significant differences were found when comparing those combinations yielding a Pearson's correlation of .53 as well as those yielding a correlation of .78.

Consider first the combinations yielding correlations of .78. As in Experiments 1 and 2, the judgments in the graphical condition with low slope, low error variance, high variance of *X* were significantly higher than judgments in the high slope, high error variance, high variance of *X* condition ($p < .05$).

Those combinations resulting in a Pearson's correlation of .53 also yielded significantly different judgments of relatedness. In the tabular format condition, the low slope, low error

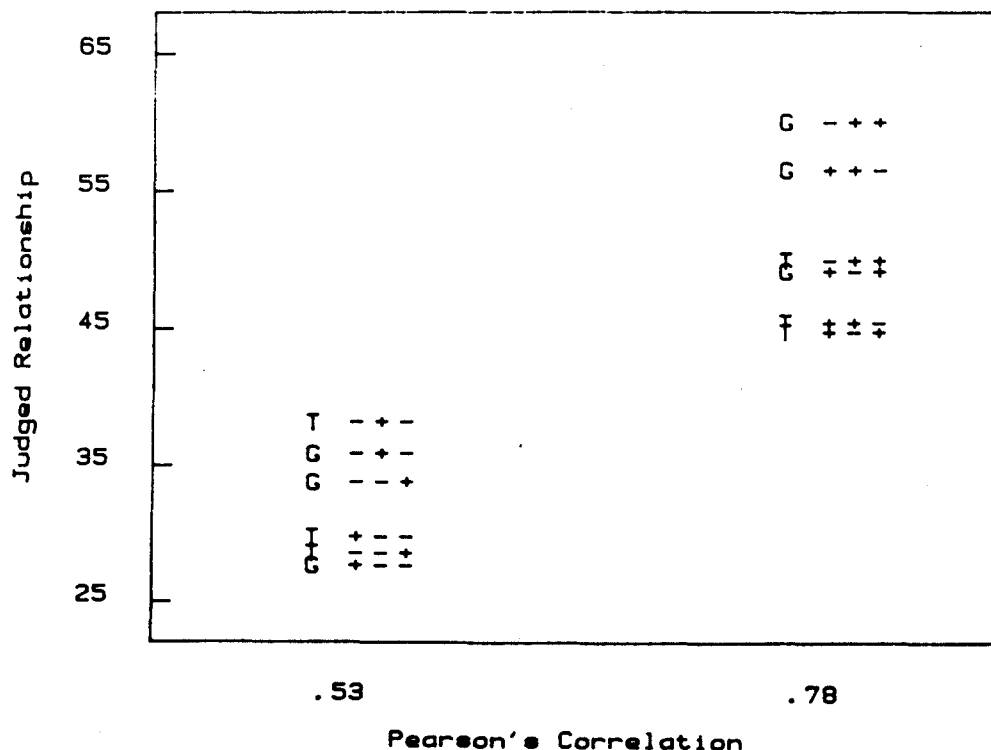


Figure 4. Mean judgments of the relation between *X* and *Y* as a function of presentation format (Graphical [G] or Tabular [T]) and slope, error, and variance of *X* of the stimulus data, Experiment 3.

variance, low variance of X combination was judged significantly higher than the low slope, high error variance, high variance of X combination ($p < .01$), indicating error variance to be a more important determinant of perceived covariation than variance of X . The former combination was also judged higher than the high slope, high error variance, low variance of X combination ($p = .027$), showing error variance to be more important than slope.

General Discussion

The major conclusion of these studies is that different ways of producing the same value of Pearson's correlation lead to different judgments of relatedness. Statistical novices and experts were more influenced by the error variance than by the variance of X or the slope. Thus, attempts to determine the psychophysical function relating values of Pearson's correlation to subjects' judgments of relatedness are neglecting relevant information. Because there are many aspects of the relation between variables, a mapping of subjects' judgments onto one measure such as Pearson's correlation can provide, at best, an oversimplified perspective on how people make judgments of relatedness.

Indeed, criteria *other* than minimizing squared deviations may be more appropriate in many situations. For instance, prediction accuracy may be defined as the proportion of times one is correct. With this as a criterion, the most important component is error variance. Even when the correlation between X and Y is low (e.g., the slope is low), the prediction of Y may be very accurate if the error variance is low. Conversely, a higher correlation may also have higher error variance (and a very high slope). From the perspective of a person operating on the environment, the most important feature of any covariation may be the accuracy with which Y is predicted, in which case the former relation is better (stronger) than the latter. Thus, as stated earlier, we do not consider the fact that subjects' judgments diverge from Pearson's r to represent a bias.

The conclusion that people are more influenced by error variance than is the Pearson procedure is consistent with previous findings on single-cue probability learning and on co-

variation detection. As discussed previously, Eade (1967) found learning was better with a low slope and a low error variance than with a high slope and a high error variance even though Pearson's r was the same. The study by Cleveland et al. (1982) was concerned not only with determining the psychophysical function between judgments and Pearson's correlations but also with the effect of the size of the cloud of points present in the scatterplot. Two scatterplots were used, the axes of which were scaled differently so that the cloud of points in one scatterplot occupied a much smaller portion of the figure than did the points in the other scatterplot. Subjects judged the degree of relatedness to be greater with the smaller cloud of points even though Pearson's correlation was the same for both. If one makes the plausible assumption that subjects based their judgments on the plot of the points without regard to the numerical labels on the axes, then the plot with the larger cloud of points had a larger variance of X and a larger error variance than did the other plot. The higher judged relatedness of the small cloud of points could thus be due to subjects' stronger weighting of error variance than of variance of X .

Other researchers have concluded that judgments are more closely related to some function of $(1 - r^2)$ than to Pearson's r (e.g., Jennings et al., 1982). This, of course, is simply the formula for the error variance with standardized variables.

It could be argued that subjects' judgments of relatedness differ for scatterplots with the same value of Pearson's correlation because subjects perceive the components inaccurately rather than because they combine these components in a different manner than they are combined in the computation of Pearson's correlation. Although we acknowledge this possibility, we suspect that subjects perceive scatterplots as integral rather than as separable stimuli and therefore do not perceive slope, error variance, and variance of X as components at all. The application of Garner's (1974) procedures for distinguishing integral from separable stimuli would seem to be a good way for future research to address this issue.

Our finding that subjects judge the relation between variables to be higher when the data are presented in a graphical format than when they are presented in a tabular format has im-

plications for one of the main conclusions drawn by Jennings et al. (1982). These authors were interested in determining the accuracy with which subjects can judge the degree of relation between real-world variables. After determining the psychophysical function relating degree of correlation to subjects' ratings, Jennings et al. were able to "translate" subjects' judgments of the degree of relation between real-world variables into Pearson correlations. Their key finding was that subjects have a tendency to overestimate the degree of relation. For example, the average judged relation between self-ratings of intellectualism and self-ratings of ambitiousness was 50 on a 100-point scale. The empirically determined value of Pearson's correlation for these two variables is .40. The rating of 50 can be translated into a correlation by determining how high a correlation must be present in hypothetical data in order for subjects to give a rating of 50. As it turns out, it took hypothetical data with a Pearson's correlation over .90 for subjects to judge the degree of relatedness to be 50. Thus, subjects would appear to overestimate the degree of relation. However, this conclusion may not be warranted. When Jennings et al. presented their hypothetical data to subjects, they were presented in a tabular format, at least in some instances. Our data indicate that presenting data in graphical format would probably have led to higher ratings; therefore, judgments of the relation between real-world variables would have been translated into lower correlations. For example, it may not have required a correlation as high as .90 for subjects to judge the degree of relatedness to be 50 if the data had been presented graphically. This is not to say that presenting the data in graphical form would necessarily have led to a more accurate translation procedure, only that Jennings et al.'s conclusions may have been dependent on their format of data presentation.

Even though our expert sample was most affected by error variance, just as the novices were, there was some indication that statistical experts do judge covariation differently. Specifically, the experts saw stronger relations in the tabular data than did the novices. However, the various interactions of format and components of the relation were similar for the dif-

ferent subject samples, suggesting that the basic covariation detection processes are the same for these two populations.

In sum, our data suggest that people can detect covariations between variables but that the components do not influence covariation judgments in precisely the same manner as they influence the value of Pearson's r . We do not wish to suggest that people perceive the three components of linear relations independently and then combine them into an overall judgment. Rather, we have been concerned with the functional relation between these components and human judgments.

References

- Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, *112*, 117-135.
- Brehmer, B. (1973). Single-cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion. *Organizational Behavior and Human Performance*, *9*, 377-395.
- Brehmer, B., & Lindberg, L. (1970). The relation between cue dependency and cue validity in single-cue probability learning with scaled cue and criterion variables. *Organizational Behavior and Human Performance*, *86*, 331-334.
- Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*, *216*, 1138-1141.
- Eade, S. (1967). *The effect of magnitude of criterion validity and positive cue validity upon human inference behavior*. Unpublished master's thesis, Ohio State University.
- Erlick, D. E., & Mills, R. G. (1967). Perceptual quantification of conditional dependency. *Journal of Experimental Psychology*, *73*, 9-14.
- Garner, W. E. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). New York: Cambridge University Press.
- Naylor, J. C., & Domine, R. K. (1981). Inferences based on uncertain data: Some experiments on the role of slope magnitude, instructions, and stimulus distribution shape on the learning of contingency relationships. *Organizational Behavior and Human Performance*, *27*, 1-31.
- Uhl, C. (1966). Effects of multiple stimulus validity and criterion dispersion on learning of interval concepts. *Journal of Experimental Psychology*, *72*, 519-527.

Received June 1, 1984

Revision received May 17, 1985 ■