Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement
2013, Vol. 45, No. 4, 313–319

0008-400X/13/$12.00 DOI: 10.1037/a0033342

What the Correlation Coefficient Really Tells Us About the Individual

Robert C. Gardner and Richard W. J. Neufeld
University of Western Ontario

The Pearson product−moment correlation is an index of how well one can predict scores on one variable given scores on another based on a linear regression equation; the higher the correlation, the better prediction. The nature of this predictability was investigated at the level of the individual. We demonstrate there is relatively little improvement in predictability of individual status for the middle 80% of the population as correlations increase from .10 to .50, and that it is not that much better for correlations as high as .90. The same is true for the middle 60% of the population when assessing how likely it is that an individual will perform as well or better (or as poorly or worse) on a second variable than on the predictor. These generalisations are demonstrated initially with an empirical sampling distribution from a population correlation of .50. It is further investigated by determining the probabilities associated with decile location based on analytical integration of the normal bivariate space for population correlations of .10, .30, .50, .70, and .90.

Keywords: conditional probabilities and correlation, correlation and predictability of individual performance, correlation from a Bayesian perspective, correlation and decile correspondence, correlation versus predictability

Considerable research is concerned with the study of individual differences, often focusing on the Pearson product–moment correlation between two variables where a significant correlation is considered an indication of some (linear) consistency in individual differences. But what does it really tell us about the individual? Cohen (1988) has characterised a correlation of .10 as depicting a small effect, commenting that "many relationships pursued in "soft" behavioural science are of this magnitude" (p. 79). He characterizes a correlation of .30 as a medium effect, observing that such values are encountered in behavioural science and that "this degree of relationship would be perceptible to the naked eye of a reasonably sensitive observer" (p. 80). He describes a correlation of .50 as a large effect, remarking that Ghiselli (1964) considered it as the practical upper limit of predictive effectiveness. These values have been described as reflecting 1%, 9%, and 25% respectively of the variance common to two variables.

There have been many articles written about the various ways in which the Pearson correlation can be interpreted and how an increase in the magnitude of the coefficient indicates an increase in predictability. For example, Rodgers and Nicewander (1988) offer 13 ways in which to interpret a correlation coefficient. In their abstract, they state "We show that Pearson's *r* (or simple functions of *r*) may variously be thought of as a special type of mean, a special type of variance, the ratio of two means, the ratio of two variances, the slope of a line, the cosine of an angle, and the tangent to an ellipse, and may be looked at from several other interesting perspectives" (p. 59). Rovine and von Eye (1997) provide a 14th way of interpreting correlation as "the proportion of times a second variable falls within a set range of the first" (p. 42), an issue that underlies the question raised here. Other articles deal with factors that influence the magnitude of the correlation coefficient (Goodwin & Leech, 2006), and the interpretation of a phi coefficient as the difference between conditional probabilities when the marginals are equal (Falk & Well, 1997). These articles deal with the interpretation of a correlation in general and not on its implications for an individual in the sample under investigation.

Our primary purpose in this investigation was to answer the question as to how well one could predict an individual's relative status on a second variable given his or her score on a first variable. A secondary aim was to answer the question of how the magnitude of the correlation between the two relate to this. Unlike most of the previous articles, our objective was to understand the implications for individual prediction, not the meaning or interpretation of *r* itself. That is, given that an individual obtains a score at the mean of the first variable, what score would you predict on the second variable? Given that an individual scored at the 10th decile on the first variable, how likely would it be for that person to score at the 10th decile of the second? And/or given that an individual scored at the *p*th decile, how likely would it be that the person would score that well or better on the second variable (or even that well or worse)? This, we propose, is the implication as far as the individual is concerned.

Robert C. Gardner and Richard W. J. Neufeld, Department of Psychology, University of Western Ontario.

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada. We thank Peter Nguyen for his diligent computer computations in generating the tables from the analytical results, and to Paul F. Tremblay for his helpful comments and recommendations on a draft.

Correspondence concerning this article should be addressed to Robert C. Gardner, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada or to Richard W. J. Neufeld, Department of Psychology, University of Western Ontario, London, Ontario N6A 3K7, Canada. E-mail: gardner@uwo.ca, rneufeld@uwo.ca

Simulation Study

Our interest in the meaning of correlation as it applies to the individual drove us to conduct an empirical study in which we investigated actual data producing a correlation of .50, corresponding to Cohen's large effect. Specifically, we obtained a sample of 1,000 observations from a population correlation of .50 using the Correlated Multivariate Random Normal Scores Generator (Aguinis, 1994). The data were in standard score form and the sample means for the two variables were −.018 and −.006, and the standard deviations were 1.000 and 1.054. The obtained correlation was .509. Figure 1 presents a scatter plot of the empirical data.

These scores were transformed to decile values, and their bivariate frequency distribution is presented in Table 1 in the form of a 10 × 10 matrix. If the correlation between the two variables were 0, this would result in a 10 × 10 matrix with an expected 10 observations in each cell. Thus, given 1,000 cases, the probability of being in any given cell is 10/1000 = .01, and the probability of being in any given decile for one of the variables is 100/1000 = .10. Correlations greater than 0 will result in some cell frequencies being greater than 10 and others less, but the marginal frequencies would still be 100 and the marginal probabilities would still be .10.

Expressed in this way, it is possible to assess (1) the probability of an observation being in any one cell, or (2) the probability of being in any one *Zy* cell or set of *Zy* cells given that one is in any one decile of *Zx*, or (3) the probability of being in any one cell or set of *Zy* cells given that one is in any set of *Zx* deciles. Returning to Table 1, consider the following examples:

1. The probability of being in any one cell is simply the frequency of that cell divided by the total *N*. Thus, the probability of simultaneously being in the 10th decile of *Zx* and the 10th decile of *Zy* is 36/1000 = .036, while the probability of being in the 5th decile of *Zx* and the 4th decile of *Zy* is 14/1000 = .014.

2. The probability of being in any one cell of *Zy* given that one is in any one decile of *Zx* is the ratio of the frequency in the particular cell divided by the total number of observations in the *Zx* decile. Thus, the probability of being in the 10th decile of *Zy* given that one is in the 10th decile of *Zx* is 36/100 = .36. Furthermore, the probability of being in any of a set of *Zy* deciles given that one is in one set of *Zx* deciles is the sum of the relevant set divided by the number of individuals in the *Zx* set. For example, the probability of being in the 8th, 9th, or 10th decile of *Zy* given one is in the 10th decile of *Zx* is (36 + 19 + 14)/100 = .69.

3. The probability of being in any one cell or set of *Zy* cells given that one is in any set of *Zx* deciles is equal to the ratio of the frequency in that cell or the sum of frequencies in the set of cells over the sum of frequencies of the *Zx* deciles. Thus, the probability of being in the 9th decile cell of *Zy* given that one is in the 7th or 8th decile of *Zx* is (10 + 11)/(100 + 100) = .105, or the probability of being in the 7th to 10th deciles of *Zy* given that one is in the 8th to 10th deciles of *Zx* is (18 + 11 + 13 + 10 + 13 + 12 + 21 + 17 + 36 + 19 + 14 + 8)/(100 + 100 + 100) = 192/300 = .64. In short, one can determine the probability of any given scenario.

This is an empirical example based on 1,000 observations drawn from a bivariate normal population with a correlation of .50, and if

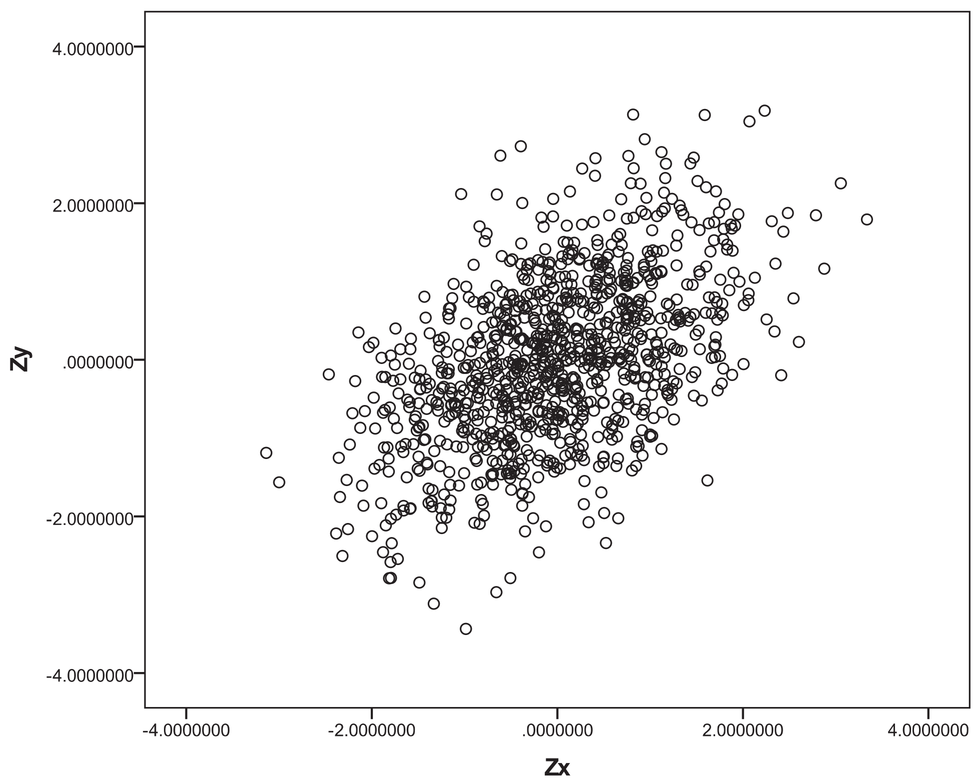

Figure 1. Scatter plot of the empirical data.

Table 1
Sample Space From a Population Correlation of $\rho = .50$

Z_Y Variable	Z_X Variable										
Decile	1	2	3	4	5	6	7	8	9	10	Row N
10	1	0	0	5	7	9	11	18	13	36	100
9	1	9	7	10	10	11	10	11	12	19	100
8	4	5	7	3	7	13	13	13	21	14	100
7	6	10	7	9	8	16	9	10	17	8	100
6	3	9	12	16	9	10	14	8	11	8	100
5	5	8	10	10	12	17	12	9	12	5	100
4	14	11	13	9	14	5	10	12	9	3	100
3	14	18	9	13	11	6	10	11	2	6	100
2	18	13	19	15	11	7	7	6	3	1	100
1	34	17	16	10	11	6	4	2	0	0	100
Column N	100	100	100	100	100	100	100	100	100	100	1000

another sample were drawn, the cell frequencies would vary slightly because of sampling fluctuations, resulting in slightly altered probabilities. A comprehensive characterization free of sampling error is provided by considering the associated probability density function for given values of the population correlation.

Theoretical Specifics

Rather than performing simulation studies to determine the conditional probabilities associated with the various deciles, the same logic can be demonstrated based on theoretical considerations.

Governing Formulae

The Pearson product–moment correlation for the population can be defined as:

$$\rho = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{N\sigma_X\sigma_Y} = \frac{\Sigma Z_X Z_Y}{N}$$

The equation for the probability density function of the bivariate normal distribution in standard score form is given by Hays (1973, p. 659) as:

$$f(Z_X, Z_Y) = \frac{1}{K}e^{-G}$$

where

$$G = \frac{(Z_X^2 + Z_Y^2 - 2\rho Z_X Z_Y)}{2(1 - \rho^2)}$$

and

$$K = 2\pi\sqrt{(1 - \rho^2)}$$

Examination of the function will reveal that it is symmetrical for any given value of the population correlation, ρ. When ρ is 0, the resulting function is a circle, but for any other value it is an ellipse with a slope given by the sign of ρ, and as the correlation increases in magnitude, the minor axis of the ellipse becomes shorter (i.e., the ellipse becomes thinner). Figure 2 illustrates the parameter space for a positive correlation between two variables, *Zx* and *Zy*. Note that it would be centered at the means of the two variables, (i.e., 0, 0).

Given the population correlation, ρ, the regression of two sets of standard scores, *Zy* on *Zx,* would have the same slope as that for *Zx* on *Zy* with a value equal to ρ, and the values of *Zx* and *Zy* would be distributed within the ellipse. The ellipse expresses the nature of the probability density function of the bivariate distribution and for any given value of ρ would be symmetrical about the centre of the ellipse. Figure 2 shows cut-off values for the *Zx* and *Zy* axes. The area designated *A* indicates the cases that are above cut-off values for both variables; the area identified as *B* indicates those that are above the *Zx* but below the *Zy* cut-offs; and the area *C* indicates those values that are below the *Zx* but above the *Zy* cut-offs. Note from the figure that the proportion of cases above the *Zx* cut-off that are also above the *Zy* cut-off is given by the ratio of *A/(A* + *B*), while the proportion of cases that are above the *Zy* cut-off that are also above the *Zx* cut-off are *A/(A* + *C*). In short, given a correlation between *Zx* and *Zy*, conditional probabilities exist which would increase with increases in the magnitude of their correlation. Thus, the Bayesian expression of the probability of exceeding Z_Y given Z_X is:

$$P(Z_Y \mid Z_X) = \frac{P(Z_Y)P(Z_X \mid Z_Y)}{P(Z_X)} = \frac{A}{A + B}$$

Similarly, the probability of exceeding Z_X given Z_Y is:

$$P(Z_X \mid Z_Y) = \frac{P(Z_X)P(Z_Y \mid Z_X)}{P(Z_Y)} = \frac{A}{A + C}$$

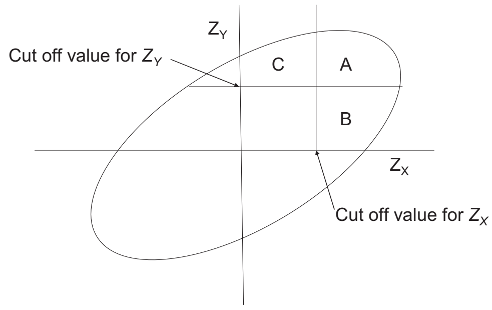

Figure 2. Parameter space of a bivariate distribution.

Note, this relationship does not depend on any cause-effect scenario. It is possible that individual differences in *Zx* are responsible for individual differences in Z*y*, or that individual differences in Z*y* are responsible for individual differences in *Zx*, or that individual differences in *Zx* and Z*y* are due to some other variable(s). The correlation model assumes simply that there is a linear association between the two variables. In research, a significant correlation implies that there exists in the population an association between the two variables, and that the unbiased estimate of the population correlation (ρ) is the value of *r* obtained in the sample. Our objective in this investigation is to demonstrate that this association has implications for the individuals sampled. The major point is that if there is a significant correlation in the population between the two variables, this means that information about one of the variables has implications about the other variable, regardless of any cause-effect interpretation. The probability density function presented above can be used to compute the areas in the intersection of the deciles in the parameter space. This is achieved by double integration of *f(Zx, Zy*) with respect to *Zx* and *Zy* over the intervals prescribed by each pair of Z_X and Z_Y deciles. That is:

$$\int_{Z_{xl}}^{Z_{xu}} \int_{Z_{yl}}^{Z_{yu}} f(Z_x, Z_y) dZ_x dZ_y$$

where Z_{xl}, Z_{xu}, Z_{yl} and Z_{yu} are the lower and upper limits of the variables *Zx* and *Zy* respectively.

Analytical Results

We investigated the bivariate probability density functions for population correlations of .10, .30, .50, .70, and .90.[1] Table 2 presents the areas (probabilities) associated with each decile in the bivariate distributions for a population correlation of .50. The cell values are given to five decimal places while the marginal values have been rounded to two places. It will be noted that each of the marginal values equals .10, and that the cell values in each row and column add to .10 with rounding, hence the sum of all the cell values equals 1.0. Being areas identified by the density function, each of these values can be viewed as the probability of membership in any given cell, while the row and column marginals are the probabilities of being in any given decile of either *Zx* or *Zy*.

Examination of the cell values will also reveal that they are perfectly symmetrical. Although it might be expected that the probabilities would be greatest for deciles in agreement between the two variables (i.e., the diagonal values), it will be noted that is not true in this table.[2]

These values can be used to estimate the probability of an individual achieving various decile cut-offs in *Zy* given a specific decile cut-off value in *Zx*. Using the values from Table 2 as examples, it can be observed that

1. the probability of being simultaneously in the 10th decile of *Zx* and the 10th decile of *Zy* is .03218;

2. The probability of being simultaneously in the 5th decile of *Zx* and the 4th decile of *Zy* is .01140;

3. the probability of being in the 10th decile of *Zx* given that one is in the 10th decile of *Zy* is .03218/.10 = .3218;

4. the probability of being in the 8th, 9th or 10th decile of *Zy* given that one is in the 10th decile of *Zx* is (.01398 + .01912 + .03218)/.10 = .06528/.10 = .6528;

5. the probability of being in the 9th decile of *Zy* given that one is in the 7th or 8th decile of *Zx* is (.01223 + .01441)/(.10 + .10) = .02664/.2 = .1332;

6. the probability of being in the 7th to 10th deciles of *Zy* given that one is in the 8th to 10th deciles of *Zx* is (.01398 + .01912 + .03218 + .01441 + .01660 + .01912 + .01369 + .01441 + .01398 + .01248 + .01223 + .01043)/(.10 + .10 + .10) = .19263/.30 = .6421.

Of course, these are expected values given that they are calculated from values defined by the density function, but it will be noted that they are very similar to those calculated from the empirical sampling distribution of the population correlation of .50 presented in Table 1.

Similar analyses can be performed for the other values of the population correlation that can be obtained from the authors. For example, it can be demonstrated that given a correlation of .10, the probability of one being in the top 50% of one distribution given that one is in the top 50% of the other is .53130, hardly a difference from the value of .50 expected in any event. The corresponding probability for each of the other correlations is .59112 for ρ = .30, .66572 for ρ = .50, .74598 for ρ = .70, and .85570 for ρ = .90. Note, therefore, that for what has been characterised as a large effect (*r* = .50) in psychological research, roughly two thirds of the sample would be correctly identified as being in the top half of the predicted distribution while one third would be misidentified. Even for a near perfect correlation (ρ = .90), only 86% are correctly placed in the top half of the distribution while 14% are misclassified.[3]

At the individual level, the correlation might be viewed as an index of how consistent an individual's position will be in both variables relevant to the other individuals in the sample, and it is assumed that the degree of consistency will increase with increases in the magnitude of the correlation. Obviously, if the correlation is 1.00, there will be perfect consistency, but what about other values of the correlation? One way of investigating this is to consider the conditional probability of an individual being in the same decile for *Zy* as for *Zx*. Table 3 presents the

[1] Tables of the bivariate normal probability distributions for correlations of .10, .30, .70, and .90 are available from the authors.

[2] It was only in the table for the correlation of .90 where this occurred. For all other tables, these probabilities are greatest only for deciles 1 and 10; for the correlation of .70, this is also true for deciles 5 and 6. In all other instances, probabilities are larger in some higher deciles in each column for those above the median and lower for those below, and the extent to which this occurs depends on the magnitude of the correlation, reflecting the width of the ellipse described by the sample space.

[3] These values may appear to be inconsistent with Rosenthal and Rubin's (1982) discussion of the binomial effect size display (BESD). Note, however, that the "*r*" values to which they refer are *phi* (φ) coefficients so that their medium effect of .30 corresponds to a Pearson correlation of .50, which is in fact a strong effect.

Table 2
Proportions of the Area for Deciles in the Population Space for a Correlation of .50

Zy	Zx										
Decile	1	2	3	4	5	6	7	8	9	10	Total
10	.00075	.00190	.00314	.00447	.00599	.00789	.01043	.01398	.01912	.03218	.10
9	.00190	.00395	.00568	.00721	.00871	.01034	.01223	.01441	.01660	.01912	.10
8	.00314	.00568	.00750	.00887	.01006	.01125	.01248	.01369	.01441	.01398	.10
7	.00447	.00721	.00887	.00992	.01071	.01140	.01204	.01248	.01223	.01043	.10
6	.00599	.00871	.01006	.01071	.01106	.01129	.01140	.01125	.01034	.00789	.10
5	.00789	.01034	.01125	.01140	.01129	.01106	.01071	.01006	.00871	.00599	.10
4	.01043	.01223	.01248	.01204	.01140	.01071	.00992	.00887	.00721	.00447	.10
3	.01398	.01441	.01369	.01248	.01125	.01006	.00887	.00750	.00568	.00314	.10
2	.01912	.01660	.01441	.01223	.01034	.00871	.00721	.00568	.00395	.00190	.10
1	.03218	.01912	.01398	.01043	.00789	.00599	.00447	.00314	.00190	.00075	.10
Total	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	1.00

probability of individuals being in corresponding deciles for each of the five correlations.

Inspection of Table 3 will reveal some important features. First, note that for each column the values in the top half are mirror images of those in the lower half. The probabilities are highest for the extreme deciles, decrease in magnitude for the more central ones, are symmetrical around the median, and become more variable as the correlation increases. Second, note that the probabilities for each decile tend to increase as the correlations increase; the mean probabilities over all deciles are .11, .14, .17, .23, and .36 for the five correlations as would be expected given that the correlations are increasing. What is probably unexpected is how low these values are. On average, the probability of an individual being in the corresponding deciles is relatively low regardless of the magnitude of the correlation though admittedly it increases with increasing correlation. On closer inspection, it will be noted that for correlations of .10, .30, and .50, there is relatively no change in predictability for deciles 2 to 9 (e.g., from .11 to .17 is the largest difference). For correlations of .70 and .90, there are slight improvements but, again, the largest changes are in deciles 1, 2, 9 and 10 (i.e., the extreme 40% of the cases).[4] Stated in another way, there is relatively little predictability with correlations from .10 to .50 for 80% of the cases, and for 60% for larger correlations.

One might argue that the purpose of correlation is not to assess the predictability of an individual being in the same decile on two variables, but rather the predictability of an individual scoring as high or higher if above the median or as low or lower if below the median. Table 4 presents the resulting probabilities for these scenarios.

As with Table 3, the values in the top part of each column are mirror images of those in the lower part. In Table 4, however, the values are lowest on the extreme ends and increase as they approach the middle, and the variability of the values in each column decreases as the correlations increase. Note, too, that like Table 3, the probabilities for each row tend to increase as the correlation increases; the mean probabilities over all rows are .39, .42, .46, .50, and .59. That is, on average, for correlations less than .70, less that 50% of the cases would be correctly identified as doing as well (or as poorly) or more extremely on *Zy* given their decile position on *Zx*. For individuals in deciles 5 or 6, there is virtually no difference in predictability, and relatively not very much more for deciles 4 and 7. That is, as in Table 3, much of the improved predictability as the correlations increase involves the extremes of the distributions, but even here the probability exceeds 50% for all deciles only when the correlation is .90.

Table 3
Conditional Probability of Co-Occurrence in the Deciles

Decile	ρ = .10	ρ = .30	ρ = .50	ρ = .70	ρ = .90
10	.13	.22	.32	.47	.69
9	.11	.14	.17	.22	.36
8	.11	.12	.14	.17	.28
7	.10	.11	.12	.15	.24
6	.10	.10	.11	.14	.22
5	.10	.10	.11	.14	.22
4	.10	.11	.12	.15	.24
3	.11	.12	.14	.17	.28
2	.11	.14	.17	.22	.36
1	.13	.22	.32	.47	.69
Mean probability	.11	.14	.17	.23	.36

Discussion

The focus of this article is the degree of correlation between two variables and the implications this has for an individual. Attention was directed to the variables in standard score form and on the individual's relative status as defined by decile membership. Although the discussion referred to positive correlations, it pertains to negative correlations as well. It is often stated that the Pearson product–moment correlation can vary from -1 to $+1$, but this is true only when the two distributions are symmetrical and have identical shapes. Two asymmetrical distributions (skewed to the left or right) with the same shape could produce a correlation of $+1$, but there would be limits on the degree of negative correlation possible. Similarly, two asymmetrical distributions that are mirror images could produce

[4] One reason for this is that with increases in the magnitude of the correlation there is a corresponding decrease in the width of the deciles, particularly in the midrange of the distributions.

Table 4
Probability of Attaining Initial or More Extreme Decile in Z_Y Given Z_X

Scenario	ρ = .10	ρ = .30	ρ = .50	ρ = .70	ρ = .90
P((9–10)9)	.23	.29	.36	.44	.59
P((8–10)8)	.33	.37	.42	.48	.58
P((7–10)7)	.41	.44	.47	.51	.58
P((6–10)6)	.50	.51	.52	.54	.59
P((5–10)5)[a]	.59	.58	.58	.58	.61
P((1–5)5)	.50	.51	.52	.54	.59
P((1–4)4	.41	.44	.47	.51	.58
P((1–3)3)	.33	.37	.42	.48	.58
P((1–2)2)	.23	.29	.36	.44	.59
Mean Probability	.39	.42	.46	.50	.59

[a] This expression refers to the probability of someone being in the 5th to the 10th decile in the predicted distribution given that they were in the 5th decile on the predictor (i.e., doing as well or better relatively in the predicted distribution as they did on the predictor), while the following one refers to the probability of someone being in the 1st to the 5th decile in the predicted distribution given that they were in the 5th decile on the predictor (i.e., doing as poorly or worse).

a correlation of −1, but there would be limits on the largest possible positive correlation that could be obtained (for algebraic considerations, see Gardner, 2000). The results presented here refer to bivariate normal distributions and apply to both positive and negative correlations, but the rationale could be generalised to any scenario provided the distributions of the two variables are known and it is assumed that the variations around the regression line are normally distributed.

Clearly, as the magnitude of the correlation increases the degree of linear relationship between two variables increases. Nonetheless, when it comes to predictability at the individual level, relatively little is gained. If the magnitude of the correlation is considered to imply a degree of predictability about the individual obtaining similar or similar and more extreme scores on both variables, then improvement of predictability holds primarily at the ends of the distributions. The view that a higher correlation implies better prediction is true only in the general sense that there is a stronger linear relationship between the two variables, but is much less so when predictions about an individual are addressed.

These results have many implications for the researcher and practitioner. Individualized prediction and associated consultation should be qualified accordingly. For example, although it is generally agreed that IQ is a strong predictor of grades in school, generally yielding a correlation of .50, generalisations that one can make about a single individual are limited. As we have seen, a statement that a student with an IQ at the mean (e.g., 100) will continue to be at the mean or better in grades is true only .58% of the time (see Table 4), while a gifted student scoring at the 9th decile in IQ will be at the 9th decile or better in grades only 36% of the time (see Table 4), and a student scoring at least at the 9th decile will have a probability of .44 of performing that well or better in grades (i.e., from Table 2 [(.01912 + .01660 + .03218 + .01912)/.20 = .44]).

These results arguably have a bearing in many contexts where the bivariate correlation is used for prediction. For example, in the clinical setting, it may be of interest to predict symptom severity on the basis of a premorbid indicator (e.g., Nicholson & Neufeld, 1993). Considering Figure 2, and related text, symptom severity in need of clinical attention may correspond to the cut-off value on the *Zy* axis, and the *Zx* axis may correspond to scores on a predictor variable. If the cut-off value on the *Zx* axis represents the point at which the prediction is that symptom severity is in need of attention, the sensitivity of measure *Zx* would comprise $A/(A + C)$, and its positive predictive power would be $A/(A + B)$. Clearly, degree of association between *Zx and Zy* has a direct bearing on such diagnostic efficiency statistics, and the bivariate probability tables can be consulted for their population values, given ρ and selected decile cut-offs.

Turning to statistical tests on single cases, a central question surrounding clinical prediction, for instance, may take the form, "Given a certain location on pretreatment variable *X*, how likely is the client's location on posttreatment variable *Y*, if there were no treatment effect?" Methods addressing this type of question engage the bivariate distribution of scores between *X* and *Y* head on (Nunnally & Kotsch, 1983; Payne & Jones, 1957). Results presented here elaborate upon the effect the degree of association plays in such statistical inference and at the level of the individual participant.

Note that the effects of ρ on the bivariate normal distribution are manifestly complex. The relation of these properties to the linear ρ itself is nonlinear. For any particular ρ, moreover, predictive efficiency is nonlinear inasmuch as it is greater at the extremes. It may be possible to develop specific functions analytically expressing these associations. However, such developments to our knowledge have yet to appear. Finally, standard practice suggests that the current observations on predictability at the level of the individual may not be widely known. We believe, however, that they represent information that every user at this level should want to know.

Résumé

Le coefficient de corrélation de Pearson est un indice qui permet de prédire les scores d'une variable sur une autre, à partir d'une équation de régression linéaire. Plus la corrélation est élevée, meilleure est la prédiction. La nature de cette prévisibilité a été étudiée au niveau individuel. Nous avons démontré qu'il y a relativement peu d'amélioration de la prévisibilité de l'état individuel pour les 80% situé au milieu de la population, alors que les corrélations augmentent de 0,10 à 0,50 pour des corrélations aussi élevées que .90, ce qui n'est guère mieux. La même chose est vraie pour les 60% du milieu de la population lorsqu'il s'agit d'évaluer la probabilité qu'une personne effectuera aussi bien ou mieux (ou aussi mal ou pire) sur une seconde variable autre que le prédicteur. Ces généralisations ont été effectuées initialement sur la base d'une distribution d'échantillonnage empirique à partir d'une corrélation de population de .50. Les probabilités associées aux déciles basés sur l'intégration analytique de l'espace normal bivariée des corrélations de la population de .10, .30, .50, .70, et ,90 ont également été examinées.

Mots-clés : probabilités conditionnelles et corrélation, corrélation et prévisibilité de la performance individuelle, corrélation à partir d'une perspective bayésienne, corrélation et correspondance décile, corrélation versus prévisibilité.

References

Aguinis, H. (1994). A quickbasic program for generating correlated multivariate random normal scores. *Educational and Psychological Measurement, 54,* 687–689. doi:10.1177/0013164494054003012

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Erlbaum.

Falk, R., & Well, A. D. (1997). Many faces of the correlation coefficient. *Journal of Statistics Education, 5,* 1–14.

Gardner, R. C. (2000). Correlation, causation, motivation and second language acquisition. *Canadian Psychology/Psychologie Canadienne, 41,* 10–24. doi:10.1037/h0086854

Ghiselli, E. E. (1964). Dr. Ghiselli comments on Dr. Tupe's note. *Personnel Psychology, 17,* 61–63.

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education, 74,* 249–266. doi:10.3200/JEXE.74.3.249-266

Hays, W. L. (1973). *Statistics for the social sciences (2nd ed.*). New York, NY: Holt, Rinehart and Winston.

Nicholson, I. R., & Neufeld, R. W. J. (1993). Classification of the schizophrenias according to symptomatology: A two-factor model. *Journal of Abnormal Psychology, 102,* 259–270. doi:10.1037/0021-843X.102.2.259

Nunnally, J. C., & Kotsch, W. E. (1983). Studies of individual subjects: Logic and methods of analysis. *British Journal of Clinical Psychology, 22,* 83–93. doi:10.1111/j.2044-8260.1983.tb00582.x

Payne, R. W., & Jones, G. J. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13,* 115–121. doi:10.1002/1097-4679(195704)13:2<115::AID-JCLP2270130203>3.0.CO;2-1

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42,* 59–66. doi:10.2307/2685263

Rosenthal, R., & Rubin, D. B. (1982). A simple, general-purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169. doi:10.1037/0022-0663.74.2.166

Rovine, M. J., & von Eye, A. (1997). A 14th way to look at a correlation coefficient: Correlation as the proportion of matches. *The American Statistician, 57,* 42–46.

Received April 24, 2012
Revision received May 7, 2013
Accepted May 9, 2013 ■