THE JOURNAL OF GENERAL PSYCHOLOGY
2019, VOL. 146, NO. 3, 325–338
https://doi.org/10.1080/00221309.2019.1585321

Common language effect size for correlations

Xiaofeng Steven Liu[a], Ryan Carlson[a], and Ken Kelley[b]

[a]University of South Carolina; [b]University of Notre Dame

ABSTRACT

The Pearson correlation coefficient can be translated to a common language effect size, which shows the probability of obtaining a certain value on one variable, given the value on the other variable. This common language effect size makes the size of a correlation coefficient understandable to laypeople. Three examples are provided to demonstrate the application of the common language effect size in interpreting Pearson correlation coefficients and multiple correlation coefficients.

ARTICLE HISTORY
Received 20 November 2017
Accepted 3 February 2019

KEYWORDS
Common language effect size; correlation

The Pearson product-moment correlation coefficient measures the linear relationship between two continuous variables. If one variable is dichotomized, the Pearson product-moment correlation becomes a point-biserial correlation. The sign of the Pearson product-moment correlation identifies the direction of the relationship between the two variables. A positive correlation implies that both variables have a tendency to increase or decrease together. A negative correlation, however, implies that when one variable increases the other has a tendency to decrease, and vice versa. The value of a Pearson correlation ranges from a negative one to a positive one, with zero meaning no association between the two variables. The absolute value of the correlation shows the strength of the relationship between the two variables. As the absolute value of the correlation increases toward one, the two variables become more related. Although the Pearson correlation is one of the most common effect sizes, it may be difficult to explain its size to non-specialists or laypeople, who can nevertheless be the stakeholders of evaluation research.

The sizes of correlations can vary widely from one field of study to another. Cohen (1988) used .1, .3, and .5 to represent "small," "medium," and "large" correlations. His rule-of-thumb numbers provide some guidance on the size of a correlation, but they are not always intelligible to laypeople, who may have trouble relating a correlation coefficient to something concrete. The difficulty with interpreting a correlation lies in the fact

CONTACT Xiaofeng Steven Liu xliu@mailbox.sc.edu Department of Educational Studies, University of South Carolina, 820 Main St, Columbia, SC 29208, USA.

that a correlation coefficient, though numerically simple, represents something abstract.

We can use the formulas to show that a correlation coefficient is indeed an abstract concept.

$$\rho = \frac{\sum_{i=1}^{n} z_x z_y}{n} \tag{1}$$

The correlation ρ is the average of the cross products of the standardized scores of the two continuous variables (z_x and z_y) among n number of subjects. If the two variables often vary in the same direction, the cross products of their standardized scores are mostly positive. So the average of those cross products is positive—a positive correlation. As the cross product does not represent any concrete entity in real life, it remains an abstract quantity. In other words, the average of the cross products (i.e., correlation) does not represent anything that people can see or feel in real life. Laypeople in particular have trouble sizing up a correlation coefficient. The problem is not unique to correlation coefficients. Laypeople have similar difficulty with Cohen's *d*, the most commonly used effect size.

McGraw and Wong (1992) suggested a common language effect size to help explain Cohen's *d* to laypeople. They used the probability of a randomly selected observation from one population being larger than a randomly selected observation from the other population. In essence, the Cohen's *d* is converted to a probability value. The idea of using probability as a way of communicating the difference between two groups is very appealing because the probability of a randomly selected observation from one group being larger over that of the other can easily resonate with people's intuition. The probability converted from Cohen's *d* serves as a bridge between the scientific findings and a broad audience. A recent study suggests that common language effect sizes are perceived as more understandable and useful than traditional effect size statistics (Brooks, Dalal, & Nolan, 2014). Therefore, the common language effect size can increase the interpretability of scientific findings and broaden their impact on the society at large. Grissom and Kim (2012) discussed the common language effect size and referred to it as the probability of superiority. Vargha and Delaney (2000) generalized the common language effect size to ordinal measures and created a general effect size, which they called the measure of stochastic superiority.

Dunlap (1994) extended the idea of common language effect size to bivariate normal correlations. A positive correlation is converted to a probability of having the same sign for the difference scores on the two variables (X and Y) between two randomly selected individuals, that is, $x_1 - x_2 > 0$ and $y_1 - y_2 > 0$ or $x_1 - x_2 < 0$ and $y_1 - y_2 < 0$. The subscripts

1 and 2 represent the two randomly selected individuals. The probability converted from the correlation can be written as

$$P[x_1 - x_2 > 0 \cap \ y_1 - y_2 > 0] + P[x_1 - x_2 < 0 \cap \ y_1 - y_2 < 0]. \tag{2}$$

It indicates how likely the order of two random individuals remains the same for the two variables X and Y. Since the above-mentioned probability expression comprises two probability terms of equal size, it can be abbreviated as

$$2P[x_1 - x_2 > 0 \cap \ y_1 - y_2 > 0]. \tag{3}$$

For a negative correlation, Dunlap suggested one minus the above-mentioned probability. For instance, the probability $2P[x_1 - x_2 > 0 \cap y_1 - y_2 > 0]$ is .63 for a positive correlation .4. If the correlation is $-.4$, the probability $2P[x_1 - x_2 > 0 \cap \ y_1 - y_2 > 0]$ will be $1-.63$ or .37. As it does not require much statistical knowledge to understand a probability value, Dunlap's common language effect size for correlation makes more sense to laypeople than a regular correlation coefficient.

In this paper, we will offer a new way to translate a correlation coefficient to a common language probability. Our approach uses probability too, but it is more flexible with the reference values on the two variables. Such flexibility allows us to tailor the interpretation of the correlations to suit different research contexts. We will demonstrate that Dunlap's joint probability can be viewed as a special case of the flexible approach. We will use a conditional probability to show how likely one variable exceeds a certain reference value, given a certain reference value on the other variable. If the two reference values on the two variables are the averages, the conditional probability mathematically equals Dunlap's joint probability for a positive correlation. We can also apply the conditional probability to interpret a multiple correlation in regression analysis because the multiple correlation is the correlation between the actual outcome and the predicted outcome.

In the following, we first discuss the conditional probability, its relation to the Pearson's correlation, and the common language effect size. We then apply the idea of common language effect size to Pearson's famous correlation of father's height and son's height. The correlation of .4 between father's height and son's height essentially means that there is a 63% probability of son's height being above the average given father's height being above the average. The correlation .4 is an abstract statistic, whereas the conditional probability 63% is the corresponding common language effect size. In addition, we provide two more examples to demonstrate the applications of common language effect size for correlation in real-world settings. One of the two examples is about military admissions, and the other example is multiple regression.

Common language effect size for Pearson correlation

Conditional probability can be used to show the relationship between two events A and B. If the two events are unrelated, then the probability of event A given event B is the same as the probability of event A, that is, $P[A|B] = P[A]$. In this case, the conditional probability, $P[A|B]$, equals the unconditional probability, $P[A]$. When the events A and B are related, the conditional probability, $P[A|B]$, will no longer be the same as the unconditional probability, $P[A]$. For instance, let us look at depression diagnosis (event A) and depression symptoms (event B). The disease occurs in the general population with a probability, $P[A]$. As the disease is related to its symptoms (event B), the probability of someone having the disease in the presence of the symptoms—the conditional probability $P[A|B]$ – is larger than the probability of having the disease in the general population. Although the value of $P[A|B]$ relative to $P[A]$ may not appear obvious, we can use Bayes theorem to explain why $P[A|B]$ is larger than $P[A]$. We can express $P[A|B]$ in terms of $P[B|A]$. By Bayes theorem, we have

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[B|A]P[A]}{P[B]} = \frac{P[B|A]}{P[B]}P[A] \tag{4}$$

The conditional probability, $P[A|B]$, is a multiple of the probability, $P[A]$. The multiplier is $P[B|A]/P[B]$. Our common sense suggests that the probability of having symptoms given the disease, $P[B|A]$, is typically higher than the probability of having symptoms, $P[B]$. The multiplier, $P[B|A]/P[B]$, is therefore larger than one. It follows that $P[A|B]$ is larger than $P[A]$.

We can use the conditional probability to show the relationship between two bivariate normal variables because their correlation can be translated to a conditional probability. We can start with the averages (μ_x and μ_y) as two reference values on the two variables for the conditional probability, because almost everyone understands the idea of an average. We will calculate the probability of the observation Y on a randomly selected individual being larger than its average μ_y, given that the observation X on that randomly selected individual exceeds its average μ_x; that is,

$$P\left[Y > \mu_y | X > \mu_x\right].$$

This conditional probability can be computed as a function of the correlation ρ between the two variables X and Y. Using Bayes theorem, we can first compute the joint probability of $Y > \mu_y$ and $X > \mu_x$ and then the conditional probability.

$$P\left[Y > \mu_y | X > \mu_x\right] = \frac{P\left[Y > \mu_y \cap X > \mu_x\right]}{P[X > \mu_x]} \tag{5}$$

The joint probability $P[Y > \mu_y \cap \; X > \mu_x]$ is the cumulative distribution function (cdf) of a bivariate normal distribution with a correlation ρ between X and Y.

We can simplify the computation by converting the observations X and Y to their respective z scores, z_x and z_y. The inequality $Y > \mu_y$ is mathematically equivalent to $z_y > 0$ because the z score of the average μ_y is zero. Likewise, the inequality $X > \mu_x$ is equivalent to $z_x > 0$. The joint probability in the numerator of Equation 5 is the cdf function of a standard bivariate normal distribution for the first quadrant, which contains all the pairs of observations with positive z scores (i.e., $z_x > 0$ and $z_y > 0$) (see Equation 6 in Appendix A).

$$P\left[Y > \mu_y \cap \; X > \mu_x\right] = P[z_y > 0 \cap \; z_x > 0] \tag{6}$$

Further, the probability $P[X > \mu_x]$ in the denominator of Equation 5 is .5 because half of the X scores are above the average μ_x. The conditional probability in Equation 5, therefore, becomes (see Equation 7 in Appendix A)

$$P\left[Y > \mu_y | X > \mu_x\right] = \frac{P\left[Y > \mu_y \cap X > \mu_x\right]}{P[X > \mu_x]} = 2P[z_y > 0 \cap \; z_x > 0] \tag{7}$$

For a negative correlation $-\rho$, the conditional probability is defined as the chance of Y below the average given X above the average,

$$P\left[Y < \mu_y | X > \mu_x\right]. \tag{8}$$

The value of the conditional probability, $P\left[Y\mu_y|X > \mu_x\right]$, will remain the same as the previous conditional probability $P\left[Y > \mu_y|X > \mu_x\right]$ for the positive correlation ρ. Unlike Dunlap's method, we use the same probability value to show the same degree of association between two variables. This makes it easier to compare the sizes of the correlations. A higher conditional probability simply means stronger association between the two variables, regardless of the sign of the correlation. The sign of the correlation, however, is reflected in the definition of the conditional probability. The users can immediately know the direction of the correlation from the expression of the conditional probability in either case.

Example 1

We can apply the conditional probability to the correlation between father's height and son's height in Pearson (1896). Let father's height be X and son's height be Y. The conditional probability, $P\left[Y > \mu_y|X > \mu_x\right]$, means the chances of having a son above the average height, given his father being above the average height. If father's height has no relationship with son's height (i.e., $\rho = 0$), the conditional probability, $P\left[Y > \mu_y|X > \mu_x\right]$, equals the unconditional probability $P[Y > \mu_y]$ or .5. In this case, there is a 50% chance that the son's height is above average. However, heredity has a bearing on a child's height. Father's height and son's heights are related—a correlation of .40. The conditional probability, $P\left[Y > \mu_y | X > \mu_x\right]$, evaluates to .63; that is,

$$P\left[Y > \mu_y | X > \mu_x\right] = 2P[z_y > 0 \cap z_x > 0] = 2 \times .3155 \approx .63.$$

We can then explain the correlation between father's height and son's height in colloquial language: There is a 63% chance that the son's height is above the average if his father's height is above the average. Thus, the relationship between father and son's height is presented in a way easily comprehensible to laypeople.

The conditional probability $P\left[Y > \mu_y | X > \mu_x\right]$ is numerically equal to the joint probability $2P[x_1 - x_2 > 0 \cap y_1 - y_2 > 0]$ for a positive correlation, although Dunlap (1994) did not explicate the idea of conditional probability in his article. The equivalence is due to the fact that the joint probability $P[x_1 - x_2 > 0 \cap y_1 - y_2 > 0]$ is mathematically equal to the cdf function, $P[z_y > 0 \cap z_x > 0]$, for a positive correlation. However, such equivalence does not carry over to negative correlations for Dunlap's joint probability.

We prefer the conditional probability approach over the joint probability approach, because the conditional probability can refer to other values than the averages μ_x and μ_y in the probabilistic expression. It is not uncommon for people to seek out a reference value other than the average on the variable because that reference value (e.g., an above-average cutoff score for competitive admission) may be desired or relevant in the specific context. The reference values for X and Y need not be the same. For example, a college admissions officer may be interested in screening applicants for entrance. He or she is concerned about the applicants' potential success in college studies, which is measured by freshman GPA (Y). If GPA 2.5 is the desired benchmark for college graduation, it makes sense to set the reference value on the variable freshman GPA to 2.5. The conditional probability approach can easily accommodate the changed reference value. We can interpret the conditional probability as the probability of having freshman GPA above 2.5, given the applicant's characteristic (X). The reference value on the X variable need not stay at the average either. It can be a different value if that value makes more sense in the context. Consequently, our approach can make the interpretation of a correlation more relevant to various research contexts.

In general, we can represent the conditional probability with any reference values (a and b) on the two variables (X and Y) as $P[Y > b | X > a]$. When computing the conditional probability, we can convert the two variables to their z scores. The conditional probability equals

$$P[Y > b|X > a] = P\left[z_y > z_b|z_x > z_a\right] = \frac{P\left[z_y > z_b \cap z_x > z_a\right]}{P[z_x > z_a]}, \tag{9}$$

where z_a and z_b are the z scores of the two reference values a and b (see Equation 9 in Appendix A). As the z score determines the percentile score, we can use percentiles to refer to z_a and z_b in our interpretation. For instance, z_a and z_b can be the 40$^{\text{th}}$ percentile and the 75$^{\text{th}}$ percentile, respectively. The conditional probability then represents the chances of someone ranking above the 75$^{\text{th}}$ percentile on Y, given his or her rank of at least 40$^{\text{th}}$ percentile on X. For a negative correlation $-\rho$, the value of the conditional probability is the same, but the probabilistic expression is different. It will be $P\left[z_y < -z_b|\ z_x > z_a\right]$ instead of $P\left[z_y > z_b|z_x > z_a\right]$.

Example 2

The Armed Services Vocational Aptitude Battery (ASVAB) test is administered by the U.S. military to determine the eligibility and qualifications of new recruits. The ASVAB test is found to be correlated with the American College Testing (ACT). The correlation between the ASVAB and ACT is .767 (Koenig, Frey, & Detterman, 2008). This finding may be of great interest to the military officers who want to find new recruits among high school graduates, who often take the ACT test. The ACT scores are accessible to the military officers and can be used to assess the eligibility of potential recruits. Unless the military officers are well versed in applied statistics, the correlation of .767 ($\rho = .767$) will read opaque to them. The common language effect size can serve as a useful tool to make the correlation comprehensible to the military officers.

Suppose that the ACT score (X) of the potential recruits at one high school is a little above the national average of the ACT score, say 22. The ACT scores have a national average of 20.9 with a standard deviation of 4.8 (i.e., $\mu_x = 20.9$ and $\sigma_x = 4.8$). The ASVAB score (Y) is usually normalized and converted to a percentile score for the purpose of recruitment. The U.S. military may use different criteria to admit new recruits to various services (e.g., U.S. Army and Air Force). For instance, Air Force recruits need to have higher percentile scores on the ASVAB than U.S. Army recruits. We can use the common language effect size for correlation to show the correspondence between an ACT score of 22 and various percentile scores of the ASVAB. Laypeople can review the correspondence and get a concrete sense of how the ASVAB and ACT scores are related.

Table 1 shows the probabilities of obtaining different percentiles on the ASVAB given the ACT score 22. To compute the conditional probability, we can first find the percentile score of the ACT 22. The percentile score is about .59. As the percentile scores correspond to unique z scores, we can

Table 1. Probability of obtaining the percentile on the ASVAB (Y) given the ACT (X) score 22.

Percentile of ASVAB (p)	$P[Y > Y_p \mid X > 22]$
.2	.98
.3	.95
.4	.90
.5	.83
.6	.73
.7	.59
.8	.43

use the z scores for the percentile scores in the actual computation. The 59^{th} percentile on the ACT 22 corresponds to a z score, $z_{.59}$. The percentile (p) on the ASVAB can vary from .01 to .99, according to the desired criterion for recruitment. The ASVAB score Y_p of $100 \times p$th percentile corresponds to its z score, z_p. The common language effect size for the correlation can be computed as a conditional probability (see Equation 10 in Appendix A):

$$P\left[Y > Y_p | X > 22\right] = P\left[z_y > z_p | z_x > z_{.59}\right]. \tag{10}$$

The probability of ranking at least 20^{th} (p = .20) on the ASVAB given the ACT 22 is evaluated to be .98. The probability of exceeding the 50^{th} percentile on the ASVAB (p = .50) given the ACT 22 is .83. If the admission criterion for elite service needs to be above the 80^{th} percentile on the ASVAB, then such probability given an ACT score of 22 is .43.

When interpreting a correlation coefficient, we can select a desired percentile (e.g., 50^{th} percentile) on the ASVAB and compare the status on the ASVAB between two different correlations (e.g., 0 vs. .767). One correlation coefficient is the obtained one (i.e., .767); and the other can be arbitrarily chosen for comparison purpose. If the chosen correlation is 0, the conditional probability, $P[Y > Y_{.5}|X > 22]$, is equal to the unconditional probability $P[Y > Y_{.5}]$, that is, $P[Y > Y_{.5}] = 1 - p = .5$. There is a 50% chance that the prospective recruit ranks above the 50^{th} percentile on the ASVAB in the presence of a zero correlation. However, the correlation between the ASVAB and the ACT is .767. The probability of ranking above the 50^{th} percentile on the ASVAB given an ACT score 22 evaluates to 83%. The improvement from 50% to 83% shows the comparison between 0 and .767 correlations, and such comparison can be easily obtained from the rows in Table 1 for other percentiles on the ASVAB.

Common language effect size for multiple correlation

The common language effect size can be used to interpret a multiple correlation coefficient in regression analysis, which assumes normality, linearity, independence, and homoscedasticity. Assume that all the assumptions of the regression analysis are tenable. The multiple correlation coefficient R is the square root of R^2, which measures how well the independent variables together predict the outcome variable Y. The model fit index, R^2, is always reported in a regression analysis, and it is also called the coefficient of determination. The R^2 can range from 0 to 1. If the R^2 is 1, the independent variables perfectly predict the outcome. A zero R^2, on the contrary, means that the predictors are useless in predicting the outcome. In practice, the R^2 falls anywhere between 0 and 1 and is a fractional number. The numerical value of R^2 appears simple, but there is no consensus on how to interpret the size of an R^2. Cohen (1988) used .0196, .13, and .26 for "small," "medium," and "large" R^2. These rule-of-thumb numbers do not convey a concrete sense of being small, medium, and large to laypeople. However, we can convert the R^2 to a common language effect size to improve its interpretability.

Our approach of common language effect size directly applies because the square root of an R^2 represents the correlation between the outcome variable Y and its predicted value $\hat{Y}$. We can use any reference value on the outcome to show how much the predictors are related to the outcome variable. We can start with the average as the reference value for simplicity of illustration. If the predictors are not related to the outcome, the conditional probability of the outcome being above the average given an above-average predicted outcome is the same as the unconditional probability of the outcome being above the average or 50%. If the predictors are related to the outcome, a multiple correlation R will be larger than zero. The probability of the outcome being larger than the average given such prediction will exceed a 50% chance. In other words, our prediction of the outcome will be better than the random guess of a coin toss, $P[Y > \mu_y | \ \hat{Y} > \mu_{\hat{y}}] > .5$ for a non-zero R. The average of the outcome is just one possible reference point, although the reference point can be taken anywhere on the scale of the variable.

We can use a series of percentile scores on the outcome in a multiple regression by way of illustration. Suppose that we regress the outcome Y on a set of predictors Xs and obtain a multiple correlation R. The predicted outcome is $\hat{Y}$. The percentile score for the outcome is denoted by a subscript p (i.e., Y_p and $\hat{Y}_p$). For instance, $\hat{Y}_{.10}$ is the 10^{th} percentile score of the predicted outcome, and $Y_{.10}$ is the 10^{th} percentile score of the actual outcome. We can change the reference value to any other point on the scale, say, the 75^{th} percentile ($Y_{.75}$). The conditional probability is then $P\left[Y > Y_{.75} \ \hat{Y} > \hat{Y}_{.75}\right]$, which means the likelihood of the actual outcome being above the 75^{th} percentile, given such prediction of the outcome from regression.

Table 2. Probability of obtaining a matching outcome given such prediction, $P[Y > Y_p|\hat{Y} > \hat{Y}_p]$.

Percentile of the outcome (p)	$P[Y > Y_p]$	$R = .14$	$R = .36$	$R = .51$
.2	.8	.81	.84	.86
.3	.7	.72	.77	.80
.4	.6	.64	.69	.73
.5	.5	.54	.62	.67
.6	.4	.45	.54	.60
.7	.3	.36	.45	.53
.8	.2	.26	.36	.44
.9	.1	.15	.25	.33

The conditional probability, $P[Y > Y_p|\hat{Y} > \hat{Y}_p]$, is a function of the multiple correlation R. The higher a multiple correlation R is, the higher the conditional probability gets. If the multiple correlation R is zero, the conditional probability, $P[Y > Y_p|\hat{Y} > \hat{Y}_p]$, is equal to the unconditional probability, $P[Y > Y_p]$. In other words, the prediction based on regression is useless in assessing the ranking of the actual outcome. For instance, the conditional probability, $P\left[Y > Y_{.75}|\hat{Y} > \hat{Y}_{.75}\right]$, will be the same as the unconditional probability $P[Y > Y_{.75}]$ or a 25% chance if the multiple correlation is zero. When the predictors are related to the outcome, the conditional probability $P\left[Y > Y_{.75}|\hat{Y} > \hat{Y}_{.75}\right]$ will exceed a 25% chance. The higher the multiple correlation R gets, the greater the difference between the conditional probability, $P[Y > Y_p|\hat{Y} > \hat{Y}_p]$ and the unconditional probability, $P[Y > Y_p]$. The different values of R will yield different conditional probabilities. Thus, we can compare one multiple correlation with another (e.g., .36 vs. .51) by reviewing the change in the conditional probability, $P[Y > Y_p\ \hat{Y} > \hat{Y}_p]$. Table 2 lists the conditional probability, $P[Y > Y_p|\hat{Y} > \hat{Y}_p]$, under three multiple correlations (i.e., .14, .36, and .51). The three multiple correlations are the square roots of Cohen's "small," "medium," and "large" R^2 : $\sqrt{.0196} = .14$, $\sqrt{.13} \cong .36$, and $\sqrt{.26} \cong .51$.

We can examine the conditional probability, $P[Y > Y_{.3}|\hat{Y} > \hat{Y}_{.3}]$, and the multiple correlation R for the 30$^{\text{th}}$ percentile of the outcome (i.e., p = .3). In the absence of any information on the predictors, the probability of a randomly selected individual ranking higher than the bottom 30% of the outcome Y is .7 (i.e., $P[Y > Y_{.3}] = .7$). Multiple regression is then used to help predict the outcome. For a multiple correlation .14 (R = .14), the regression explains the outcome to a limited extent. If the predicted outcome is believed to rank higher than the bottom 30% of the predicted outcome ($\hat{Y} > \hat{Y}_{.3}$), the probability of the actual outcome ranking higher than the bottom 30% given this prediction is .72 ($P\left[Y > Y_{.3}|\hat{Y} > \hat{Y}_{.3}\right] = .72$). Obviously, the odds of ranking higher than the bottom 30% given such prediction is not much different from the odds without using any prediction. This is because the predictors together do not explain a great amount of the variation in the outcome—the multiple correlation is low (R = .14). However, the conditional probability will improve if the predictors as a whole explain the outcome well (R = .51). In this case, the conditional probability, $P\left[Y > Y_{.3} | \hat{Y} > \hat{Y}_{.3}\right]$, will rise to .80. Similarly, we can explain any other percentile score of the outcome with reference to different multiple correlations.

Example 3

We can use the common language effect size to explain how well a multiple regression helps select employees for satisfactory job performance. Suppose that there is a job performance criterion on a normalized scale (Y) and a few continuous predictors of job performance. A multiple regression can be used to make prediction on job performance, given the information on the predictors. If regression is not used to predict job performance, 50% of potential employees will naturally score above the average of the job performance criterion. Without using predictors and regression, the chances that the hired employees exceed the average of the job criterion are basically a coin toss. However, if a regression analysis is used to select employees, based on the predicted above-average performance, then the chances of the employees' outperforming the average given such a prediction will exceed 50%. In other words, the odds start to move in favor of above-average job performance. The 50% chance is the unconditional probability without the help of any predictors and regression, whereas the better than 50% chance is the conditional probability of above-average job performance, given such a prediction from the regression, that is, $P\left[Y > Y_{.5} | \hat{Y} > \hat{Y}_{.5}\right]$. The conditional probability conceptually resembles the "proportion satisfactory among those selected" in Taylor and Russell (1939), which is used in lieu of a correlation between the selection test and job performance in the studies of personnel selection. The higher the multiple correlation R the regression produces, the higher the conditional probability, $P\left[Y > Y_{.5} | \hat{Y} > \hat{Y}_{.5}\right]$, will become. We can examine this conditional probabilities for Cohen's "small," "medium," and "large" multiple correlations (i.e., .14, .36, and .51). The conditional probability, $P\left[Y > Y_{.5} | \hat{Y} > \hat{Y}_{.5}\right]$, evaluates to .54, .62, and .67 for Cohen's "small," "medium," and "large" multiple correlations, respectively (see the fourth row in Table 2). In effect, these probabilities allow us to recalibrate the multiple correlation coefficients and render them intelligible to laypeople. For instance, if we interpret a multiple correlation .51 in a traditional way, we can say that the actual job performance and the predicted job performance have a correlation of .51. This explanation is not very intuitive to non-statisticians. Using the common language effect size, we can explain that the multiple correlation .51 means a .67 probability of above-average job performance given such a prediction from regression. Laypeople can then discern the size of the multiple correlation in a sensible way.

Conclusion

There has been an increasing emphasis on reporting effect sizes in psychological research. The Pearson correlation coefficient is one of the most common effect sizes. Despite its numerical simplicity, it requires some statistical sophistication to fully appreciate the size of a correlation coefficient, which may pose a challenge to non-specialists or laypeople. They can nevertheless be the stakeholders of psychological testing. The standard effect sizes often work for specialists and statisticians but not necessarily for laypeople. As the old adage suggests, one size does not fit all. Non-traditional effect sizes like common language effect sizes can be used to suit the consumers of psychological research who are not well trained in statistics (Brooks et al., 2014).

The common language effect size for correlation offers an alternative way to view the relationship between two bivariate normal variables. If there is a correlation between the two variables, knowing the values on one variable allows us to assess the probability of obtaining certain values on the other variable. This is a conditional probability, and it has often been used to shed light on one event with information on another related event. In a similar way, a correlation coefficient—be it a Pearson correlation or a multiple correlation—can be converted to a conditional probability to show how one variable is related to the other. The probabilistic expression is flexible in referring to different reference values on the two variables, and it can be tailored to suit various research contexts.

The common language effect size for correlation shares the same characteristic as the previously published common language effect sizes because they are all expressed in probability (Dunlap, 1994; McGraw & Wong, 1992; Vargha & Delaney, 2000). It makes a correlation coefficient comprehensible to laypeople, who are not knowledgeable about statistics but are consumers of scientific research. The common language effect size enables a researcher to interpret a correlation coefficient to non-statisticians in colloquial probability terms.

References

Brooks, M. E., Dalal, D. K., & Nolan, K. P. (2014). Are common language effect sizes easier to understand than traditional effect sizes? *Journal of Applied Psychology*, *99*(2), 332–340. doi:10.1037/a0034745

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

Dunlap, W. P. (1994). Generalizing the common language effect size indicator to bivariate normal correlations. *Psychological Bulletin*, *116*(3), 509–511. doi:10.1037/0033-2909.116.3.509

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications*. New York, NY: Routledge.

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence*, *36*(2), 153–160. doi:10.1016/j.intell.2007.03.005

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365. doi:10.1037//0033-2909.111.2.361

Pearson, K. (1896). Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London*, *187*(0), 253–318. doi:10.1098/rsta.1896.0007

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, *23*(5), 565–578. doi:10.1037/h0057079

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*, 101–132. doi:10.2307/1165329

APPENDIX A: COMPUTATIONAL FORMULAS

Equation 6:

$$P\left[Y > \mu_y \cap \ X > \mu_x\right] = P[z_y > 0 \cap \ z_x > 0] = \int_0^{+\infty}\int_0^{+\infty} p(z_x, z_y, \rho)dz_x dz_y,$$

where $p(z_x, z_y, \rho)$ is the probability density function of a standard bivariate normal distribution with a correlation ρ between the two variables. The double integral can be easily obtained with the help of statistical software.

Equation 7:

$$P\left[Y > \mu_y | X \ > \ \mu_x\right] = \frac{P\left[Y > \mu_y \cap \ X > \mu_x\right]}{P[X \ > \ \mu_x]} = 2P[z_y > 0 \cap \ z_x > 0] = 2\int_0^{+\infty}\int_0^{+\infty} p(z_x, z_y, \rho)dz_x dz_y$$

Equation 9:

$$P[Y > b | X \ > \ a] = P\left[z_y > z_b | z_x \ > \ z_a\right] \ = \frac{P[z_y > z_b \ \cap \ z_x > z_a]}{P[z_x > z_a]} \ = \frac{1}{P \ [z_x \ > \ z_a]}\int_{z_b}^{+\infty}\int_{z_a}^{+\infty} p(z_x, z_y, \rho)dz_x dz_y$$

Equation 10:

$$P\left[Y \ > \ Y_p | X \ > \ 22\right] = P\left[z_y > z_p | \ z_x \ > \ z_{.59}\right] = \frac{1}{1 - .59}\int_{z_p}^{+\infty}\int_{z_{.59}}^{+\infty} p(z_x, z_y, \rho)dz_x dz_y$$

APPENDIX B: R CODE

```
library(mvtnorm)

  # example 1
  2*pmvnorm(lower = c(0,0), upper = c(+Inf,+Inf), mean = c(0,0),
     corr = matrix(c(1,.4, .4,1),ncol = 2)) [[1]]
  #Table 1 in example 2
  for(i in 2:8){
  p = i/10
  zb = qnorm(p)
  za=(22-20.9)/4.8
  pxy = pmvnorm(lower = c(zb,za), upper = c(+Inf,+Inf), mean = c(0,0),
     corr = matrix(c(1,.767,.767,1),ncol = 2))[[1]]
  py_x = pxy/(1-pnorm(za))
  cat("\t", p,"P[y > b|x > a]=",py_x,"\n")
  }
  # last column of Table 2 in example 3
  for(i in 2:9){
     p = i/10
     zp = qnorm(p)
     pxy = pmvnorm(lower = c(zp,zp), upper = c(+Inf,+Inf), mean = c(0,0),
     corr = matrix(c(1,.51,.51,1),ncol = 2))[[1]]
     py_x = pxy/(1-p)
     cat(p,"\t"," P[y > y_p|yhat > yhat_p]=",py_x,"\n")
  }
```