I'm Less and Less Happy Until Finally I'm Sad: Estimating Correlations When Variables Divide a Construct into Parts

Kimberly A. Barchard
University of Nevada, Las Vegas

James A. Russell
Boston College

UNLV

BOSTON COLLEGE

The Problem

   Testing whether two variables – such as happiness and sadness – are best represented as separate dimensions or as opposite parts of a single bipolar dimension is controversial. Imagine that happiness and sadness are two variables that form a single bipolar dimension. When that bipolar dimension is measured as a whole (with response options from 1 = *very happy* to 5 = *very sad*), it correlates -1 with its reverse (which uses response options 1 = *very sad* to 5 = *very happy*). As the first variable increases, the second always decreases. However, when the bipolar dimension is divided into two parts that are measured separately, their correlation is far from -1. Imagine we ask participants, "How happy are you?" (response options 0 = *Not at all* to 10 = *Extremely*), and, "How sad are you?" (same response options). If happiness and sadness are mutually exclusive parts of a single bipolar construct, then any degree of happiness implies a complete lack of sadness and vice versa: Each variable is 0 whenever the other is positive. See Figure 1. This makes it impossible for the variables to correlate -1. For example, if these two variables divide their bipolar dimension exactly in half, and if scores on the full bipolar dimension have a normal distribution, these variables correlate -.467 (Russell & Carroll, 1999).

   Cacioppo and Bernstein (1994) noted that we cannot test if a variable is bipolar by using bipolar items: Measuring happy-sad as a single bipolar variable presupposes its bipolarity. Therefore, they argued that researchers need to divide potentially bipolar variables into two parts and measure each part separately. Still, researchers mistakenly assumed that the correlation between the two parts should be approximately -1. When they found correlations near -.467, they concluded the two were not parts of a single bipolar dimension. We invented a method to estimate the correlation between the whole constructs based upon data from measures that capture only parts of those constructs.

When exact opposites divide a construct into two parts, they do not correlate -1. We estimate the correlation between the constructs when measures do not cover the whole constructs.

Figure 1
Opposites that each measure a part of a bipolar dimension
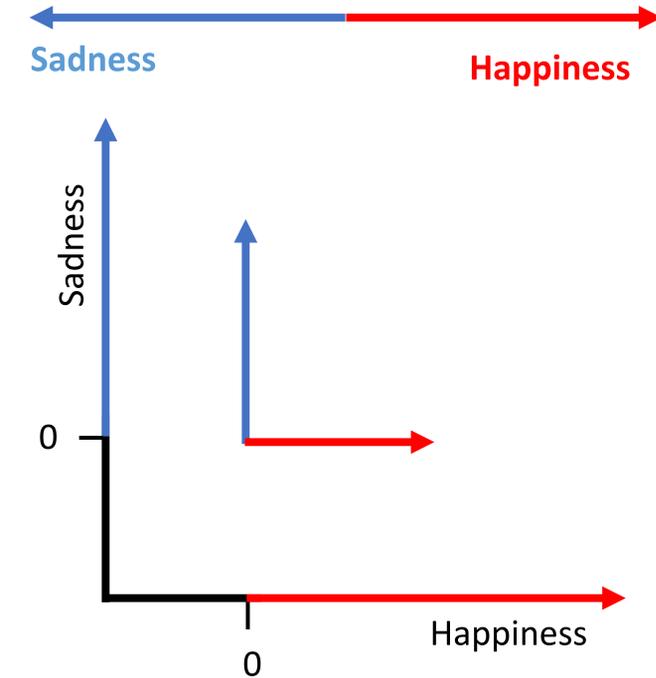
The Solution: Simulation

   First, calculate the correlation between the two observed variables, x and y, and then calculate a 95% confidence interval for the population correlation, ρ_{xy}. A convenient confidence interval calculator is given at http://vassarstats.net/rho.html   For example, imagine that r_{xy} = -.320 in a sample of 400 participants. This website calculator shows that a 95% confidence interval for ρ_{xy} is given by the interval [-.405, -.230].

   Second, correct the end points of the confidence interval for lack of reliability to create a confidence interval for the correlation between perfectly reliable variables. Use the correction for attenuation formula (Spearman, 1904). For example, imagine that the reliability of x is estimated as .70 and the reliability of y is estimated as .60. In this case, a 95% confidence interval for the correlation between perfectly reliable x and y, $\rho_{xy,corrected}$, is given by the interval [-.625, -.355].

   Third, estimate the proportions of the constructs covered by each variable. For example, if 30% of scores on happiness are 0, researchers estimate x covers 70% of X (and that the remaining scores were censored). If 60% of scores on sadness are 0, researchers estimates y covers 40% of Y (and the other 60% were censored). Enter these proportions into our Excel file.

   Fourth, select a null hypothesis to test. If we want to test whether happiness (X) and sadness (Y) are opposites, we would hypothesize that $\rho_{XY} = -1$. Enter this number into our Excel file.

   Fifth, the Excel file randomly generates 10,000 cases (or 500,000 cases) for which X and Y have the specified correlation and the specified proportions of censoring, and which are also normally distributed. Excel calculates r_{xy} for those 10,000 cases. For example, if we hypothesize $\rho_{XY} = -1$ and indicate that x covers 70% of X, and y covers 40% of Y, Excel calculates r_{xy} = -.538.

   Fifth, compare the Excel-generated correlation (for which the null hypothesis is true) with the 95% confidence interval calculated above, [-.625, -.355]. Because -.538 falls inside this confidence interval, we retain the null hypothesis $\rho_{XY} = -1$.

   Sixth, to create a confidence interval for ρ_{XY}, vary the values of ρ_{XY} in Excel to find r_{xy} values that match the end points of the confidence interval generated in step 2. Using this approach, we found that a 95% confidence interval for ρ_{XY} is given by the interval [-1, -.55].

The Details

Visit http://barchard.faculty.unlv.edu/examining-opposites/ to obtain the Excel files and a detailed handout for this poster, which describes each step in more detail.