The attenuation of correlation coefficients: a statistical literacy issue

David Trafimow
New Mexico State University
e-mail: dtrafimo@nmsu.edu

Summary Much of the science reported in the media depends on correlation coefficients. But the size of correlation coefficients depends, in part, on the reliability with which the correlated variables are measured. Understanding this is a statistical literacy issue.

Keywords: Attenuation; Correction for attenuation; Correlation coefficients; Regression to the mean; Reliability.

INTRODUCTION

Much of the research to which young people are exposed comes from the media and is in the form of correlations that are presented in as dramatic a fashion as possible so as to entice viewers. An important and well-known issue for statistical literacy is that correlation does not mean causation. Young people that are statistically literate, in this respect, should be less willing to jump to the conclusion that one of the correlated variables causes the other. In addition, the realization that the correlation coefficient might not be causal carries with it the possibility of making the correlation coefficient seem less important than it otherwise would seem.

Another statistical literacy issue pertains to the size of correlation coefficients. Specifically, the correlation coefficients that researchers obtain, and that are reported in the media, likely are attenuated by the measures not being perfectly reliable. If one corrects for such attenuation, the magnitude of the 'corrected' correlation coefficient likely will be larger than the obtained one, suggesting that it might actually be more important than it otherwise would seem to be. My goal is to discuss this issue of the attenuation of correlation coefficients.

The attenuation effect was first discovered by Spearman (1904). Although Spearman provided a lengthy mathematical proof of the effect and how to correct for it, the effect can be understood intuitively as an example of the well-known phenomenon of regression to the mean, which is important for students to understand in its own right. To understand regression to the mean, it is useful to consider an extreme example of complete randomness. Imagine that there are five pieces of paper, and each paper has a number written on it, with the numbers being 1, 2, 3, 4 and 5 for each of them, respectively. Suppose that each of five people receive one of the pieces of paper at random. The mean score will be 3. However, someone will receive a score of 5, and someone will receive a score of 1. Suppose that the pieces of paper are collected and then randomly distributed again to the five people; the best guess of what any person will score is 3 because this guess minimizes the possibility of error. To see that this is so intuitively, consider that if one guesses 3, the worst error possible is to be two points off if the person receives a 1 or a 5. In contrast, if one guesses 2 or 4, it is possible to be off by three points if the person receives a 5 or 1, respectively. And if one guesses 1 or 5, it is possible to be off by four points if the person receives a 5 or 1, respectively. In summary, even if a person received an extreme score, such as a 1 or 5 on the first trial, the best guess for even the extreme person's score on the next trial is the mean, which equals 3.

Or suppose that pieces of paper are distributed on the second trial according to the rule that whatever people received on the first trial is what they receive on the second trial. In that case, of course, the best guess – and you will be right every time – is whatever the person received on the previous trial. In this case, there is no randomness, and so, there is no regression to the mean. In most of the sciences, including economics, psychology, medicine and many others, the data are somewhere between no randomness whatsoever and complete randomness.

To move in the direction of the attenuation of correlation coefficients due to unreliability of the measures, consider yet another example where a person takes a knowledge test and obtains a score that is much higher than the mean. Why did the person acquire such a high score? There are two answers. One answer is that the person actually has a high degree of knowledge. The other answer is that the person happened to be particularly lucky, as would become obvious if that person took the test additional times and received lower scores. On average, for practically all tests, both answers will be true to a greater or lesser extent; that is, scores are due partly to the person's knowledge and partly to randomness. In general, as we noted with the example of people receiving pieces of paper with numbers written on them, more randomness implies more regression to the mean and less randomness implies less regression to the mean. Thus, if we knew how much randomness influences scores on the test, this knowledge would provide us with a strong clue about how seriously we should take extreme scores. Less randomness would imply that we should take an extreme score more seriously, whereas more randomness would imply that we should take an extreme score less seriously because of the expectation of substantial regression to the mean.

And so, we finally come to the issue of reliability of tests. Although reliability of tests is an extremely complicated topic and there are multiple ways of indexing a test's reliability, the general idea can be explained rather easily. In essence, the idea is that if a test is reliable and people take the test twice, they should obtain approximately the same score on both test-taking occasions. Assuming that a sample of people, rather than a single person, takes the test twice, we would expect scores on the two test-taking occasions to correlate with each other; this correlation coefficient is one way to measure the reliability of the test. Clearly, to the extent to which people's scores on the test are due to randomness, we would expect the reliability to be low, whereas we would expect the reliability to be high if there is very little randomness. Thus, reliability can be considered to be an inverse measure of randomness. The greater the reliability of the test, the less randomness plays into the scores, whereas the lower the reliability of the test, the more randomness plays into the scores. Thus, in terms of regression to the mean, more reliability implies less regression to the mean and less reliability implies more regression to the mean.

But what if there are two variables? It should be obvious from the foregoing discussion that for both variables, luck plays less of a role when reliability is high and luck plays more of a role when reliability is low. For an extreme example, imagine that one flips two coins and correlates heads or tails on one coin with heads or tails on the other coin. Coin flips have zero reliability, and so, the expected correlation also will be zero. Put more generally, as reliabilities decrease and luck plays more of a factor, the magnitudes of correlation coefficients are pushed in the direction of zero (positive correlations become less positive and negative correlations become less negative). This is the famous attenuation effect whereby the magnitudes of obtained correlation coefficients are lower (closer to zero) than they would be if the two variables were measured with perfect reliability.

Equation (1) is often termed the 'attenuation formula' and summarizes the foregoing qualitative discussion in a quantitative way, where r_{XY} is the observed correlation between measures of two constructs (validity), $reliability_X$ is the reliability of the measure of construct X, $reliability_Y$ is the reliability of the measure of construct Y and R_{XY} is the 'true' correlation or the correlation that would be expected if the measures were perfectly reliable.

$$r_{XY} = R_{XY}\sqrt{reliability_X reliability_Y} \tag{1}$$

Equation (1) shows how the correlation a researcher is likely to obtain is lower than the correlation that the researcher could potentially obtain if only the measures of the constructs were perfectly reliable. For many purposes, it is useful to estimate this true correlation (R_{XY}), and the estimation is facilitated by a simple algebraic manipulation of equation (1), to obtain equation (2) below. Equation (2) is the 'dis-attenuation' formula.

$$R_{XY} = \frac{r_{XY}}{\sqrt{reliability_X reliability_Y}} \tag{2}$$

Consider an illustrative example. In psychology, 0.7 is often considered to be the unofficial dividing line between levels of acceptable reliability and unacceptable reliability. Suppose that the true correlation between two variables is 0.8 and the reliabilities of the measures of both variables are at the barely acceptable level of 0.7. The expected correlation between the two variables would be $0.8\sqrt{(0.7)\ (0.7)} = 0.56$, which is much less than the true 0.8 level. Going the other way (using equation (2) instead of equation (1)), if a student is presented with an obtained correlation coefficient of 0.56 and with the reliabilities of the individual measures, the implication is that the true correlation is 0.8.

The teacher who wishes students to understand that reported correlations could dramatically underestimate true correlations might have difficulty making the case by using equation (1) or equation

(2). Although equations (1) and (2) are easy for the teacher to put on the chalkboard or on a projector screen, many students have difficulty visualizing what the equations really mean. For example, although students tend to understand that an implication of equation (1) is that observed correlations tend to be lower than true correlations, they fail to appreciate the extent of the effect or its importance. The difficulty students have in visualizing the implications of mathematical equations has been noted elsewhere. For example, Trafimow (2011) showed how to improve students' understanding of Bayes' theorem by providing them with a proof by pictures. In the case of the attenuation effect, the problem students have is in understanding just how much of a problem unreliable measures pose for the researcher who wishes to obtain a respectable observed correlation. Thus, my goal is to enable students to be able to better perceive this attenuation effect. To that end, I propose figure 1.

To create figure 1 in a way that is easy for students to understand, it was necessary to simplify equation (1). If we define p as the product of the reliabilities of the measures ($p = reliability_X\, reliability_Y$), equation (1) simplifies to equation (3) below.

$$r_{XY} = R_{XY}\sqrt{p} \tag{3}$$

Figure 1 takes advantage of the simplicity of equation (3). The horizontal axis represents the product of the reliabilities (p), whereas the vertical axis represents the correlation a researcher can expect to observe (r_{XY}). The curves represent the cases where the true correlation (R_{XY}) is 0.2 (bottom curve), 0.4, 0.6, 0.8 or 1 (top curve). The student can see at a glance, merely by following the curves from the right to the left (as in many Semitic languages), how the unreliability of the measures attenuates the observed correlation relative to the true correlation. My experience has been that students have an 'ah hah' experience when presented with figure 1. The attenuation effect is no longer a mathematical abstraction but something tangible that they can actually see. In terms of a popular cliché, 'Seeing is believing'.

Finally, to really drive in the lesson, it is interesting to consider examples that have received attention from the media. Robertson et al. (2013) reported a correlation of 0.18 between TV viewing and aggression; more TV viewing led to more aggression. Unfortunately, the researchers did not report the reliabilities of their TV viewing or aggression measures, but an advantage of the example is that the popular press rarely reports reliabilities anyhow (even the original articles often do not report reliabilities), and so, students might as well understand the ambiguity of such reports from the start. To continue, however, we might imagine that the reliabilities have various values. For instance, we might imagine that the reliability of the TV viewing measure is 0.7 and that the reliability of the aggression measure is 0.8, in which case equation (2) suggests that the best estimate of the true correlation is $\frac{0.18}{\sqrt{(0.7)(0.8)}} = 0.24$. Alternatively, we might imagine that these reliabilities are 0.5 and 0.6, respectively, in which case the best estimate of the true correlation is $\frac{0.18}{\sqrt{(0.5)(0.6)}} = 0.33$. Well, then, although a correlation of 0.18 might be considered to be rather unimpressive, 0.24 might be considered to be an improvement, and 0.33 might considered to be quite impressive, at least relative to 0.18. It might be an interesting intellectual exercise for students to generate reasons why reliability might be impressive or

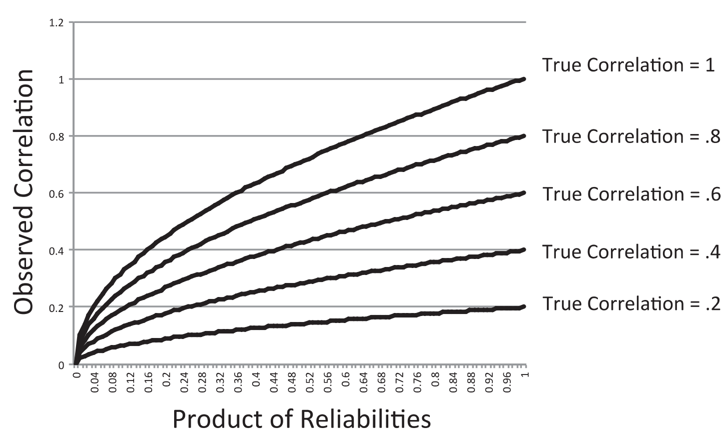

Fig. 1. The observed correlation coefficient is expressed as a function of the product of the reliabilities of the two measures and the true correlation coefficient

unimpressive and to use equation (2) to work out the implications of this intellectual exercise for what the true correlation might be.

For a second example, let us consider an article by Carpenter (2012) that received extensive attention by the media. An important finding is a correlation of 0.27 between a component of narcissism – grandiose exhibitionism (EE) – and self-promotion on Facebook. Carpenter reported the reliabilities of both measures, and these were 0.83 for EE and 0.84 for self-promotion. Note that these are impressive reliabilities, and so, we would expect less of an effect when dis-attenuating than we saw in the foregoing example. Again employing equation (2), we have the following estimate of the true correlation: $\frac{0.27}{\sqrt{(0.83)(0.84)}} = 0.32$. Even with Carpenter's relatively impressive reliabilities, the use of equation (2) nevertheless resulted in an improvement of $0.32 - 0.27 = 0.05$, going from the obtained correlation to the best estimate of the true correlation.

In conclusion, correlations can be small, large or somewhere in between, and the size matters. Students should be trained the following: (a) to ask what the size of the correlation is in the first place and (b) to understand how the unreliability of measures attenuates correlation sizes. A dis-attenuated correlation might differ considerably from that which is reported, and students need to be cognizant of this likely possibility. To be an intelligent consumer of correlational findings reported in the media, it is necessary for students to have a strong appreciation of the effects of the reliabilities of the measures on correlation sizes.

References

Carpenter, C.J. (2012). Narcissism on Facebook: self-promotional and anti-social behavior. *Personality and Individual Differences*, **52**(4), 482–486.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72–101.

Trafimow, D. (2011). Using pictures to enhance students' understanding of Bayes' theorem. *Teaching Statistics*, **33**, 83–84.

Robertson, L.A., McAnally, H.M., and Hancox, R.J. (2013). Childhood and adolescent television viewing and antisocial behavior in early adulthood. *Pediatrics*, **131**(3), 439–446.