Psychological Methods
2018, Vol. 23, No. 4, 690–707

1082-989X/18/$12.00 http://dx.doi.org/10.1037/met0000153

Sidelining the Mean: The Relative Variability Index as a Generic Mean-Corrected Variability Measure for Bounded Variables

Merijn Mestdagh, Madeline Pe, Wiebe Pestman, Stijn Verdonck, Peter Kuppens, and Francis Tuerlinckx
University of Leuven

Abstract

Variability indices are a key measure of interest across diverse fields, in and outside psychology. A crucial problem for any research relying on variability measures however is that variability is severely confounded with the mean, especially when measurements are bounded, which is often the case in psychology (e.g., participants are asked "rate how happy you feel now between 0 and 100?"). While a number of solutions to this problem have been proposed, none of these are sufficient or generic. As a result, conclusions on the basis of research relying on variability measures may be unjustified. Here, we introduce a generic solution to this problem by proposing a relative variability index that is not confounded with the mean by taking into account the maximum possible variance given an observed mean. The proposed index is studied theoretically and we offer an analytical solution for the proposed index. Associated software tools (in R and MATLAB) have been developed to compute the relative index for measures of standard deviation, relative range, relative interquartile distance and relative root mean squared successive difference. In five data examples, we show how the relative variability index solves the problem of confound with the mean, and document how the use of the relative variability measure can lead to different conclusions, compared with when conventional variability measures are used. Among others, we show that the variability of negative emotions, a core feature of patients with borderline disorder, may be an effect solely driven by the mean of these negative emotions.

Translational Abstract

The variability of processes is important across diverse fields, in and outside psychology. When measurements of these processes are bounded, which is often the case in psychology (e.g., participants are asked "rate how happy you feel now between 0 and 100?"), most variability indices become confounded with the mean. This is problematic for interpreting findings related to variability (effects of manipulation, correlations with other variables), as it is unclear whether they truly reflect effects involving variability, or are just a side effect of the mean. In the worst case, conclusions on the basis of research relying on existing variability measures may be unjustified. Here, we introduce a generic solution to this problem by proposing a relative variability index that is not confounded with the mean. The proposed index is studied theoretically and we offer an analytical solution for the proposed index, along with software tools (in R and MATLAB) to compute the relative index for measures of standard deviation, relative range, relative interquartile distance and relative root mean squared successive difference. In five data examples, we show how the relative variability index solves the problem of confound with the mean, and document how the use of the relative variability measure can lead to different conclusions, compared with when conventional variability measures are used. Among others, we show that the variability of negative emotions, a core feature of patients with borderline disorder, may be an effect solely driven by the mean of these negative emotions.

Keywords: bounded data, variability, confounded with the mean

Supplemental materials: http://dx.doi.org/10.1037/met0000153.supp

This article was published Online First April 12, 2018.

Merijn Mestdagh, Madeline Pe, Wiebe Pestman, Stijn Verdonck, Peter Kuppens, and Francis Tuerlinckx, Faculty of Psychology and Educational Sciences, University of Leuven.

The research leading to the results reported in this paper was sponsored in part by the Fund of Scientific Research Flanders by the grant for Merijn Mestdagh, in part by the Belgian Federal Science Policy within the framework of the Interuniversity Attraction Poles program (IAP/P7/06) and in part by the Grant GOA/15/003 from the KU Leuven. No data were gathered for this article. All data used in the article were previously discussed in other articles. For each data set we clearly refer to this article. The general idea was first presented as a poster at the IOPS (Interuniversity Graduate School for Psychometrics and Sociometrics) Winter conference (2014) and later in a 15-min presentation at the IMPS (International Meeting of the Psychometric Society) conference in Beijing (2015).

Correspondence concerning this article should be addressed to Merijn Mestdagh, Faculty of Psychology and Educational Sciences, University of Leuven, Tiensestraat 102, Box 3713, 3000 Leuven. E-mail: merijn.mestdagh@kuleuven.be

Most psychological characteristics are not fixed or immutable. On the contrary, they often tend to change and fluctuate across time, observers, objects, and contexts (Fleeson & Law, 2015; Nesselroade & Salthouse, 2004). Our emotions fluctuate with the ebb and flow of daily life, our behavior continuously adjusts to the circumstances we find ourselves in, and so on. Even our personality, values, and attitudes, attributes long considered to form the basis of our stable sense of self, are found to interact with internal and external factors causing their expression to fluctuate with time (Baird, Le, & Lucas, 2006; Fleeson, 2004; Wittenbrink, Judd, & Park, 2001). The patterns of these within-person changes are studied as an integral part of psychology (Molenaar & Campbell, 2009), with the magnitude of such changes being regarded as essential in explaining or understanding various psychological phenomena (Diehl, Hooker, & Sliwinski, 2014). For example, variability in emotion and mood is believed to play a large role in well-being (Houben, Van den Noortgate, & Kuppens, 2015), variability in behavior is thought essential to understand the nature of personality (Fleeson & Wilt, 2010), variability in cognition is regarded as an indicator of impending cognitive decline or low functionality (Ram, Rabbitt, Stollery, & Nesselroade, 2005), variability in physiology like heart-rate is considered a prime indicator of parasympathic activity indicative of regulatory capacity (Koval, Ogrinz et al., 2013; Segerstrom & Nes, 2007) and so on. In sum, aside from simply examining single instances or trait-levels of psychological attributes (as reflected by the average levels of feelings, behavior, and cognition), researchers are increasingly studying intraindividual variability across a large variety of domains. The study of variability is not limited to within-person variability as is exemplified by the study of, for instance, income inequality (Piketty, 2014), or happiness variability over individuals within a certain country (Kalmijn & Veenhoven, 2005).

Despite the large and growing interest in measures of (intraindividual) variability, it remains marred with a fundamental problem: Variability is closely intertwined with the mean, especially when the measurements are bounded. Bounded measurement scales are very common in psychology (e.g., numerical scales from 0 to 100). In such bounded measurements, there is an inherent structural relation: Depending on the mean, the range of possible variability scores is limited differently.[1] Consequently, the variability of a variable is strongly confounded with the mean as a consequence of the measurement instrument and this confound may then result in erroneous conclusions. For instance, one may incorrectly conclude that the set of intraindividual variability scores (e.g., within-person standard deviation in negative affect) and some external trait variable (e.g., depression) correlate, while this correlation may in fact be entirely due to its confound with the mean.

This problem has been long known (Baird et al., 2006; Eid & Diener, 1999; Kalmijn & Veenhoven, 2005), and corrections for the confounding problem have been proposed. Yet, these corrections are problematic, can only be used in very specific situations or require sophisticated analysis methods (an overview of these corrections will be given below). In this article, we address this conundrum by offering an alternative yet simple solution: the relative variability index, which is defined as the ratio of the variability divided by the maximum possible variability given the mean. This correction method removes the inherent structural relation between the intraindividual mean and variability for any bounded measure, resulting in a variability measure that is not confounded by the mean.

In this article, we will focus our attention mostly on the standard deviation as a measure of variability. However, we also offer equivalent solutions for several other variability measures such as the range, the interquartile range, and the root mean squared successive difference (Jahng, Wood, & Trull, 2008). In addition, we would also like to point out that while in our writing we focus on within-person indices of variability, the outlined problem and solution hold in fact for any measure of variability (e.g., within-country variability).

The remainder of the article is organized as follows. First, we formally define the various variability measures. Second, we demonstrate why and how these variability measures are confounded with the mean. Third, we discuss why existing proposals that deal with this confound are limited or problematic. We then propose an alternative solution that eliminates this confound labeled the relative variability index and study it theoretically. Finally, we illustrate the use of the relative index in a number of (large) real-life data sets, and demonstrate in two data sets how one can arrive at different conclusions depending on whether the confound is properly taken into account or not.

Method

Measures of Intraindividual Variability

Over the years, intraindividual variability has been operationalized in different ways. In order to introduce these measures, let us first give some notation. Consider a participant i (where i can range from 1 to K). Assume N_i measures are collected from each participant (e.g., repeated assessment of an emotion item). Based on this set of measurements for participant i, we can compute the average M_i and a variability measure V_i. M_i is given by

$$M_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{i,j}$$

where $x_{i,j}$ is measurement j of individual i.

As for a variability measure V_i, there are many options. The simplest and most prominent of the many alternative operationalizations is the intraindividual standard deviation (*SD*; Nesselroade & Salthouse, 2004; Ram & Gerstorf, 2009). The *SD* is a function of the average (squared) deviation observed within-person fluctuations display around their mean level:

$$SD_i = \sqrt{\frac{1}{N_i - 1} \sum_{j=1}^{N_i} (x_{i,j} - M_i)^2}.$$

The use of the intraindividual *SD* to characterize within-person change is long-standing and widespread, and for several good reasons: It captures a key aspect of the time-dynamics of a process or attribute (namely the amplitude of its changes), it is relatively

[1] Obviously, this can be formulated the other way around: Depending on the variability, the range of possible mean scores is limited differently. However, in this article, we consider the mean to be the first and primary statistic to consider. This means that for variability to play a role, it should contribute something over and above the mean.

simple to calculate, and, it is the most commonly used measure of spread.

Other measures have been used as well, such as the root mean squared successive difference (*RMSSD*) which is a function of the averaged squared change of an individual from one time point to the next (Jahng et al., 2008):

$$RMSSD_i = \sqrt{\frac{1}{N_i - 1}\sum_{j=1}^{N_i - 1}(x_{i,j+1} - x_{i,j})^2}.$$

Yet other measures reflecting variability for person i one may consider are the range (R_i, the difference between the maximum and minimum observation of a person) or the interquartile range (IQR_i, defined as the difference between the third and first quartile).

In this article, we take it for granted that calculating averages and standard deviations (or RMSSD, etc.) on a sample of observations are meaningful operations. If we measure the sample of shirt number for a footballer throughout his career, then an average and standard deviation are not very meaningful numbers. However, for a set of ratings to an emotion item (e.g., how happy are you?), such calculations are meaningful and we follow a long tradition in psychology here. In addition, a variability measure such as RMSSD requires even more from the design to make sense: In case of the RMSSD, the data need to have an underlying time ordering. This issue will be revisited in the Discussion.

The Problem: Confounding With the Mean

In practice, we find that when measurements are bounded, this variability measure V is related to the mean M (when we refer in general to the mean and variability measures, the person index i is dropped). As a consequence of the relation between V and M, results regarding V will be confounded by M (Baird et al., 2006; Eid & Diener, 1999; Kalmijn & Veenhoven, 2005). To explain this confounding, consider a measurement of subjective momentary happiness bounded between 0 and 100, for example, because participants were asked "How happy do you feel now (between 0 and 100)?" Obviously, the mean will be bounded between 0 and 100. But more interestingly, the variability will also be bounded. Variability measures are always bounded from below by zero (i.e., they are always positive), but for bounded measurements the variability will also have a maximum. This maximal attainable value for the variability measure is dependent on the mean. However, the exact functional form (i.e., how the maximum attainable variability varies as a function of the mean) is far less obvious. Due to this maximum of the variability measure, and because it is a function of the mean, the variability measure itself will also be related to the mean.

Let us elaborate our example one step further. Assume a collection of 1,000 (simulated) time series of each six measurements (thus $N_i = 6$ for all i). All measurements are bounded between 0 and 100. Four instances of these time series are shown in the top panel of Figure 1. In the four middle panels, the relation between the mean and four variability measures (*SD*, *RMSSD*, *IQR*, and *R*, respectively) is shown for all 1,000 time series. As expected, all the means are bounded between 0 and 100 and all variability measures are bounded from below by zero.

The maximum possible variability, denoted by $\max(V|M)$ is shown by a red line. The precise functional form of $\max(V|M)$ depends on the particular variability measure used. As can be seen from Figure 1, the type of dependence can be quite complex. As the maximum variability is dependent on the mean, the variability measure itself will also be related to the mean. In an extreme case, the mean of the time series even defines the variability exactly. For example, if the measurements are bounded between 0 and 100, the time series of a person i with mean $M_i = 100$ always implies a variability measure $V_i = 0$. This is easy to see: The only way a set of values can have such a mean is if all measurements are equal to 100, therefore the variability must be zero.

The direct consequence of the relation between the mean and the variability is that, when evaluating the effect of a manipulation on, or the association between a variable of interest D and the variability measure V, it is quite likely that the average M is a third variable that may (partly) account for the observed effect or relation.

The Existing Solutions and Why They are Problematic

Broadly speaking, one can distinguish between four types of existing solutions to the problem of the relation between mean and variability for bounded measurements that can be easily implemented.

First, the easiest and most common way (e.g., Koval, Pe, Meers, & Kuppens, 2013) to tackle this problem is to use regression analyses and statistically control for the mean by including it as a predictor variable:

$$D_i = \beta_0 + \beta_1 V_i + \beta_2 M_i + \varepsilon_i. \quad (1)$$

Such an approach ensures that the variability V is not credited for the linear relation between the mean M and the outcome variable D.

However, the use of this regression solution is not without problems. First, in some situations the strong dependence between M and V may lead to problems of multicollinearity. Second, nonlinear dependencies between M and D are ignored. Third, whatever dependency between M and V exist, the relation between another variable of interest and the variability V will be difficult to interpret. For example, β_1 is commonly understood as the amount that D is expected to increase if V is increased by one and M is held constant. However, if M and V are related to each other, it makes no sense to assume V will increase while M stays constant, making the interpretation problematic. This problem becomes worse when there is multicollinearity.

A second proposed solution is to create a flexible parametric or nonparametric model to relate V to M: $V_i = \hat{f}(M_i) + \varepsilon_i$ (Baird et al., 2006; Eid & Diener, 1999). We can then use the residual $\varepsilon_i = V_i - \hat{f}(M_i)$ as a purified substitute for V_i. Again, there are several disadvantages with this approach. First of all, $V_i - \hat{f}(M)$ is difficult to interpret as a variability measure. Second, we only want to undo the effects of the bounds but we do not want to partial out any meaningful variation in V. However, if both V and M jointly depend on another psychological construct, both V and M would also be related to each other (Kalmijn & Veenhoven, 2005) and $\hat{f}(M)$ would be strongly influenced by this extra interdependence. Third, if $\hat{f}(\cdot)$ is chosen to be a nonlinear function, the result can become very sensitive to noise in the data. A third solution is to use the coefficient of variation:

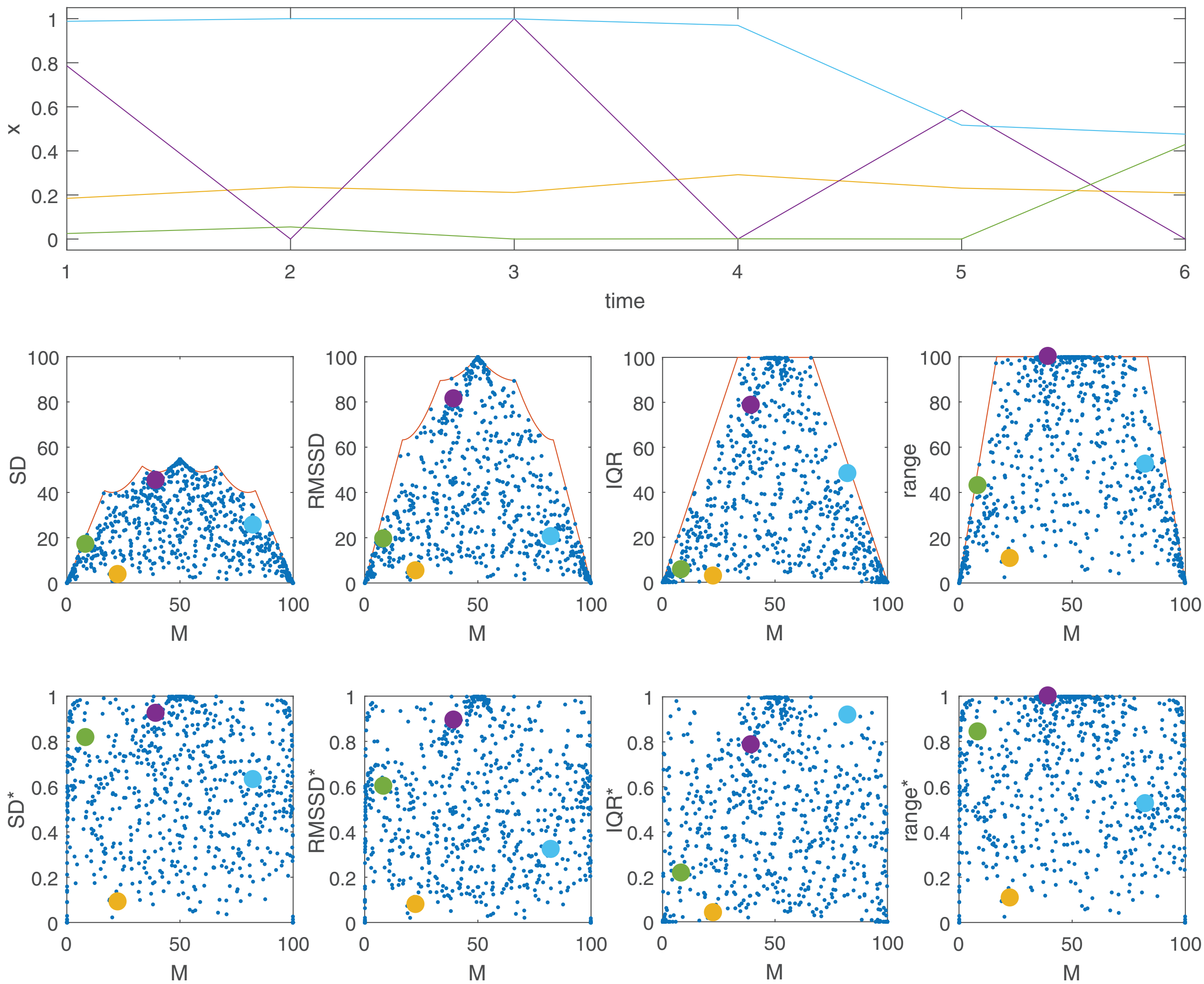

Figure 1. This figure shows the influence of the bounds on the relation between the mean and various variability measures. We simulated 1,000 random time series with each six time points ($N_i = 6$ for all i) using a beta autoregressive model (Rocha & Cribari-Neto, 2008) and random parameters for the coefficients (the mean and the inverse of the precision were sampled from a uniform distribution between 0 and 1, the autoregressive parameter was sampled from a uniform distribution between −1 and 1). In the top panel four such time series are shown. For each time series the mean and four variability measures are computed and these are shown in the four middle panels, with the variability measures V plotted against the mean M. The points in these scatterplots originating from the top four time series are shown in the corresponding color. $\max(V|M)$ is plotted as a red line. On the bottom four panels, the relative variability measures V^* are shown against the mean M. From left to right the variability measures are the *SD*, the *RMSSD*, the *IQR*, and the *range*. Again, the four colored dots are the dots associated with the time series shown in the top panel.

$$CV_i = \frac{SD_i}{M_i}. \quad (2)$$

Unfortunately, the *CV* also suffers from two shortcomings. First, it can only be used when variability is measured with the standard deviation; there is no analogue for other variability measures, such as the *RMSSD* or the range *R*. Second, it is designed only for a situation where the data are bounded by zero from below (e.g., reaction time [RT] or income), but not by an upper limit.

A fourth and last solution makes use of latent variable models (Lesaffre, Rizopoulos, & Tsonaka, 2007; Skrondal & Rabe-Hesketh, 2004). Several approaches are possible. A first approach is proposed by Lesaffre, Rizopoulos, and Tsonaka (2007) and based on the logit normal distribution, which is a bounded distribution. A random variable X follows a logit normal distribution if the logit transform of X (i.e., the inverse mapping function given by $log\left(\frac{X-L}{U-X}\right)$, with L and U being the lower and upper boundary, respectively) follows a normal distribution with mean M_{latent} and standard deviation with SD_{latent}

(Aitchison & Shen, 1980). Lesaffre et al. (2007) adjusted the model so that it can be applied to discrete data (e.g., all integers between 0 and 100, as is used in various experience sampling [ESM] data scales) but also such that the boundary values can be observed. In a first step, the bounded distribution is divided into bins so that each discrete point is in the middle of one bin (e.g., for the observation 50, the bin ranges from 49.5 to 50.5). Then the likelihood of each discrete point is defined as the probability mass of the logit transform of this bin, according to the latent normal distribution. For example data point $x_{i,j} = 50$, included in the bin ranging from 49.5 to 50.5, of participant i has the following contribution to the likelihood:

$$L(x_{i,j} = 50) = \Phi\left(\frac{logit(50.5) - M_{latent,i}}{SD_{latent,i}}\right) - \Phi\left(\frac{logit(49.5) - M_{latent,i}}{SD_{latent,i}}\right), \quad (3)$$

where Φ is the cumulative distribution function of the standard normal distribution. The product of the individual likelihoods of all the data points can then be numerically maximized to estimate $M_{latent,i}$ and $SD_{latent,i}$.

A second approach to make use of a latent variable model for bounded measurements assumes an unbounded distribution (e.g., a normal or a logistic) underlying the measurements of each participant but in this case the end points of the scale acts as thresholds: the probability mass falling below and above the thresholds in the latent model is then considered to be the probability of observing the lower and upper endpoint, respectively.[2] Such models are related to Tobit regression models (Skrondal & Rabe-Hesketh, 2004). We will not further consider this second approach but only include the first approach proposed by Lesaffre et al. (2007), but the obstacles to applying the latter approach that are given below, also apply to the extension of Tobit regression.

In both examples, both the mean and the variance of the latent distribution are assumed to be person specific. As these (person specific) latent distributions are unbounded for the mean and only bounded from below for the variance they do not constrain one another, and so no dependency is induced. There are three drawbacks of this solution. First, all the methods rely on latent distributions which are parametrized using the mean and the standard deviation or variance. It is unclear how other variability measures such as the range *R* and the *RMSSD* should be incorporated. Second, this solution also gives a new interpretation to the person specific means (e.g., the latent mean can be smaller than zero while all observations are larger than zero) and so the person specific mean cannot be computed anymore using the sample average. This would make the use of such methods backward incompatible with the existing extensive literature on the effect of the mean. It is not known how the sample average, which is most used in the literature, could be related to the same constructs as the latent mean. Third, the solution may lead to quite complex models that have to be estimated per participant. This may be hard because of insufficient data. A solution is then to work with a mixed or hierarchical model, but then a computationally intense numerical estimation process is required (Hedeker, Berbaum, & Mermelstein, 2006; Hedeker, Demirtas, & Mermelstein, 2009). In addition, to evaluate the association between a variability measure and variable of interest D in a hierarchical model, one must allow for covariates influencing the person specific variability of these latent distributions (Hedeker et al., 2006, 2009; Lesaffre et al., 2007). In both cases (hierarchical or not), the models are nonstandard and this makes these approaches less suited for the applied researcher.

To sum up, we find ourselves in a situation in which researchers want to study within-person variability and its antecedents, correlates, and consequences, but the measures used to study variability are severely confounded with the mean. Several solutions have been proposed, but each of the currently proposed solutions has a number of shortcomings. As a result, most studies in this field gravitate toward the simplest solutions: using linear regression.

The Relative Variability Index

We propose the relative variability index as an alternative solution to the problem that measures of variability are confounded with the mean when relying on bounded measurements. Our solution can be applied widely to several variability measures such as the *SD*, *RMSSD*, *IQR*, and *R*.

The fundamental cause of the dependency between the mean and variability is the shape of the region of support of the mean and the variability measures (the region of support is defined as the region where mean-variability observations can occur). For all variability measures, shown in the middle row plots of Figure 1, the region of support has a nonrectangular shape. This means that the mean and variability are not independent variables. More technically, if we assume that the mean and the variability are continuous random variables[3] and the region of support is not the Cartesian product of the univariate (or marginal) regions of support, the mean and variability cannot be considered independent random variables (see Holland & Wang, 1986).

Counteracting the dependence. To rectify the dependence between mean and variability, we will define the relative variability index V^* that transforms the bivariate region of support into a rectangle. The easiest way to achieve this is to divide by the upper bound:

$$V_i^* = \frac{V_i}{\max(V_i|M_i)}, \quad (4)$$

where $\max(V_i|M_i)$ is the maximum variability given a mean M_i for individual i. The relative variability is denoted with a star, for example, SD^* for the relative standard deviation or $RMSSD^*$ for the relative *RMSSD*.

It is evident from the definition that V^* is restricted to lie between zero and one: The relative variability can be seen as the proportion of variability that is observed, relative to the maximum possible variability that can be observed given a certain mean.

[2] Again we assume that our measurements are on a continuous scale and for which calculating, for example, a sample average is a meaningful operation. In case of purely discrete ordinal data, there are a number of thresholds such that the areas under the latent distribution below, between, and above thresholds corresponds to the probabilities of responding in the categories. However, the disadvantages mentioned below also apply to this approach.

[3] If the random variable representing an individual measurement is continuous, then the mean and variability will be continuous as well. This assumption may also hold approximately when using a fine-grained measurement scale (e.g., from zero to 100) or when having a coarser scale but averaging across many measurements (i.e., a large k_i).

After the correction, V^* is much less confounded by the mean. This is shown in the lower panels of Figure 1. The relative variability tries to extract the unique information, independent of the mean.

The meaningfulness of the relative variability can further be illustrated with hypothetical data where a participant is repeatedly measured on a scale between 0 and 100. A participant with an alternating time series of zero and one has the same standard deviation as a participant with alternating ratings of 50 and 51. However, the first participant will have a much larger relative standard deviation. We think that this difference in relative standard deviation is justified because the difference between zero and one (with an average of 0.5) is intuitively perceived larger than the difference between 50 and 51 (with an average of 50.5).

The correction from Equation 3 will not ensure that V^* and M are by default independent because the relative variability only removes the dependence that was introduced by the boundedness of the scale. Of course, both the variability and the mean may depend jointly on another psychological construct, which would also result in interdependence (Kalmijn & Veenhoven, 2005).

For the *SD* (and variance), *RMSSD* (and *MSSD*), *IQR*, and *R* (i.e., the range), we derived analytical expressions that makes it possible to easily compute $\max(V|M)$; (see online supplemental material; Kružík, 2000). Some of the calculations (e.g., for $RMSSD^*$) rely on an algorithm to find the exact solution, but no numerical optimization is needed, which avoids a large computational burden. Therefore, calculating V^* is straightforward. Our methods are implemented in MATLAB and R, and they are available online (http://ppw.kuleuven.be/okp/software/relative_variability/).

Behavior close to the bounds and weighting. If M is exactly equal to the lower or upper bound, no variability is observed because each individual observation that goes into the computation exactly equals M. In this case, $\max(V|M)$ will be zero and therefore the division in Equation 3 is not possible. Participants with a mean value M equal to one of the bounds, will have to be omitted from further analysis.

However, measuring the variability for participants with a M close to the bounds is also not straightforward. One could compare the current method with a magnifying glass, making it still possible to detect variability in the data near the bounds. In an ideal world where measurements had infinite precision and no measurement error, this would not be of major concern. Unfortunately, that is not the case. As a result, our magnifying glass (i.e., the relative variability index) may produce an undesirable amplification of errors in these regions. Our relative variability index shares this problem with the more known coefficient of variation, where a similar mechanism is at play close to the zero lower bound.

Luckily, we know exactly by how much the errors near the bounds are inflated, namely by one over the maximum possible variability. This means that the error variance of the relative variability is inflated with one over this maximum squared. Therefore, we can take the inflation into account in any further analysis, for example by weighting each relative variability with the inverse of this inflation factor,

$$w_i = (\max(V_i|M_i))^2. \tag{5}$$

Of course, it is impossible to know whether persons with data near the bounds really have a low uncorrected variability (e.g., *SD*) or whether the diminished variability is solely due to the bounds. In any case, no analysis should be too dependent on these participants with extreme means and variabilities. Therefore, also if one chooses not to use the relative variability one may check the influence of these persons by using a weighted analysis (e.g., weighted regression; Faraway, 2004) and investigate if the analysis is sensitive for this change.

Prior to the use of the weights of Equation 5 in an analysis, the weights should be normalized so that they sum to the number of observations in the analysis: $\sum_{i=1}^{K} w_i = K$ (K being the number of participants). In an unweighted linear regression, such as Equation 1, the parameters are found by minimizing the sum of squared errors $\sum_{i=1}^{K} \varepsilon_i^2$. In the weighted variant with K observations the parameters are estimated by minimizing

$$\sum_{i=1}^{K} w_i \varepsilon_i^2,$$

the weighted sum of squared errors.[4]

This weighted linear regression is designed for data where the variance of ε_i is inversely related with w_i (Faraway, 2004), which is then reflected in the criterion variable. In our case however we only know the increased variance of V^*, one of the predictors in the model. In this situation, using a weighting scheme leads to a decreased effective sample size, denoted as K_{eff}. A useful formula for K_{eff} can be found in important sampling (Elvira, Martino, Luengo, & Bugallo, 2017; Kong, 1992):

$$K_{\text{eff}} = \frac{K^2}{\sum_{i=1}^{K} w_i^2}. \tag{6}$$

K_{eff} is maximal and equal to K when all the weights are equal, and normalized to $w_i = 1$. A major consequence of weighting is that the degrees of freedom of the analysis should be adapted to represent the K_{eff}. If not all the weights are the same, K_{eff} will be smaller than K, leading to a decrease in power. Because K_{eff} may be a real number, the degrees of freedom in the regression analyses may be real numbers (much as in repeated measures ANOVA where a nonsphericity correction is applied).

Using the weighting scheme from Equation 5, results can be interpreted as the product of any linear regression. As in regression analysis, it is not recommended to extrapolate and to overgeneralize the interpretation of the results to situations for which there were no data. Similarly, it is not recommended to make predictions for new observations with means near the bounds. Even if there were observations near the bounds with which the regression was build, they were significantly down-weighted.

The Relative Standard Deviation and Its Relation to Other Variability Measures

If the relative variability is computed for the standard deviation (i.e., SD^*) there are some interesting relations with existing indices. In fact, the relative standard deviation can be seen as an extension of the (independently developed) ζ coefficient (Golay, Fagot, & Lecerf, 2013):

[4] In both MATLAB and R, one can assign weights for any generalized linear model using `glmfit` and `glm`, respectively (MATLAB, 2016; R Core Team, 2015). Examples on how to use the relative variability in combination with a weighted linear regression are given in the online package in both MATLAB and R.

$$\zeta_i = \frac{SD}{\lim_{N_i \to \infty} \max(SD_i | M_i)} \approx SD^*.$$

If the number of measurements per participant N_i goes to infinity and the standard deviation is used as variability measure, the relative variability index gives similar results as the ζ coefficient.

Although the ζ coefficient is based on an analogue idea (dividing the standard deviation by its limiting maximum value), it has some shortcomings: It is not suited for small samples, it is only developed to use for the standard deviation of accuracy measures, and it cannot be applied to other variability measures. In addition, the index has never been tested on real data.

Second, if the measurements are only bounded by a lower limit (e.g., 0) without an upper bound, then the relative standard deviation SD^* is proportional to the coefficient of variation (see online supplemental material). This means that the SD^* will result in the same conclusion as the coefficient of variation in data sets where every person has the same number of measurements. Thus, the SD^* can be considered as a generalization of the coefficient of variation for doubly bounded measures.

The Relative Standard Deviation for Some Common Bounded Distributions

To get a deeper understanding of the relative variability index, we will study it in the context of some known (and lesser known) distributions defined with a bounded region of support. As there are often analytical results for the standard deviation of these distributions we will focus here on the relative standard deviation. Because most bounded distributions are originally defined between zero and one (and every measurement instrument can be rescaled to this interval), also we will adopt these bounds in the following paragraphs.

Bernoulli distribution. The most basic bounded distribution is probably the Bernoulli distribution leading to a binary outcome where the only possible outcomes are the bounds themselves. An example time series would be a series of $N_i = 6$ correct ($x_{i,j} = 1$) or wrong ($x_{i,j} = 0$) answers (e.g., {1, 0, 0, 1, 0, 1}).

The relation between the mean M and the standard deviation SD (calculated based on repeatedly observed binary random variables) for different persons with different probabilities p_i is shown in Figure 2. For the Bernoulli, the standard deviation is not only bounded by the mean, but it is an exact function of the mean (see online supplemental material):

$$SD_i = \sqrt{M_i(1 - M_i)\frac{N_i}{N_i - 1}}.$$

Any relation we find between the standard deviation and an outcome variable D is just an indirect effect of the mean. As shown in the online supplemental material, the relative standard deviation reveals that there is indeed no extra information in the variability. The relative standard deviation is exactly equal to one for all persons (irrespective of their observed series or their mean):

$$SD_i^* = 1.$$

Binomial distribution. Next, we look at the binomial distribution. As the binomial distribution is a generalization of the Bernoulli model, it will lead to similar results. Assume that a single observation for a participant is generated by a binomial distribution with number of trials n and success probability p_i. Then the probability of observing a proportion of m successes in n trials is given by:

$$\Pr\left(x_{i,j} = \frac{m}{n}\right) = \binom{n}{m} p_i^m (1 - p_i)^{n-m}. \quad (7)$$

An example series could consist out of the grades on $N_i = 6$ subsequent tests, each scored on $n = 10$ points, such as $\left\{\frac{8}{10}, \frac{7}{10}, \frac{6}{10}, \frac{10}{10}, \frac{9}{10}, \frac{8}{10}\right\} = \{0.8, 0.7, 0.6, 1, 0.9, 0.8\}$. Assuming the length of the time series N_i goes to infinity, the population mean and standard deviation are given by

$$M_i = p_i$$

and

$$SD_i = \sqrt{\frac{M_i(1 - M_i)}{n}}.$$

Again there is a clear relation between the mean and the standard deviation. As opposed to the normal standard deviation, the relative standard deviation is only a function of the number of trials n:

$$SD_i^* = \frac{1}{\sqrt{n}},$$

again showing that standard deviation does not add anything beyond the mean. Using simulations we show in Figure 2 that the results above also hold approximately for N_i much smaller than infinity.

Beta distribution. Another well-known bounded distribution is the beta distribution. A common parametrization of the beta distribution uses the mean μ and the concentration or precision v as parameters (Kruschke, 2015; Rocha & Cribari-Neto, 2008). The concentration v is a measure of how concentrated the distribution is around a certain mean μ. As shown in Figure 2 and in the online supplemental material, sampling from distributions with the same concentration will lead to the same relative standard deviation. It is also interesting that due to the inflation problem near the bounds, the relative standard deviation becomes much more variable. This is clearly illustrated in Figure 2.

Logit normal distribution. As a last distribution we will discuss is the logit normal distribution (Aitchison & Shen, 1980), which has been used by Lesaffre et al. (2007) as discussed above. This distribution can be constructed by an inverse logit transform of the normal (which maps an unbounded random variable to the unit interval). The parameters of the logit normal are the mean μ_{latent} and standard deviation σ_{latent} of the underlying normal. Unfortunately, the logit normal has no analytical expressions for the mean and standard deviation (on the bounded interval). But as is shown in the simulations in Figure 2, the relative standard deviation is nearly constant for a large latent standard deviation. For a lower value of the latent standard deviation, the relative standard deviation flattens the relation the between mean and variability, but it seems to undercorrect (i.e., there is still some degree of dependence). Still, the relative standard deviation is much more similar to the σ_{latent} as the normal standard deviation which does not correct at all: In Figure 2, over all simulated latent standard deviations the correlation between σ_{latent} and the SD^* is

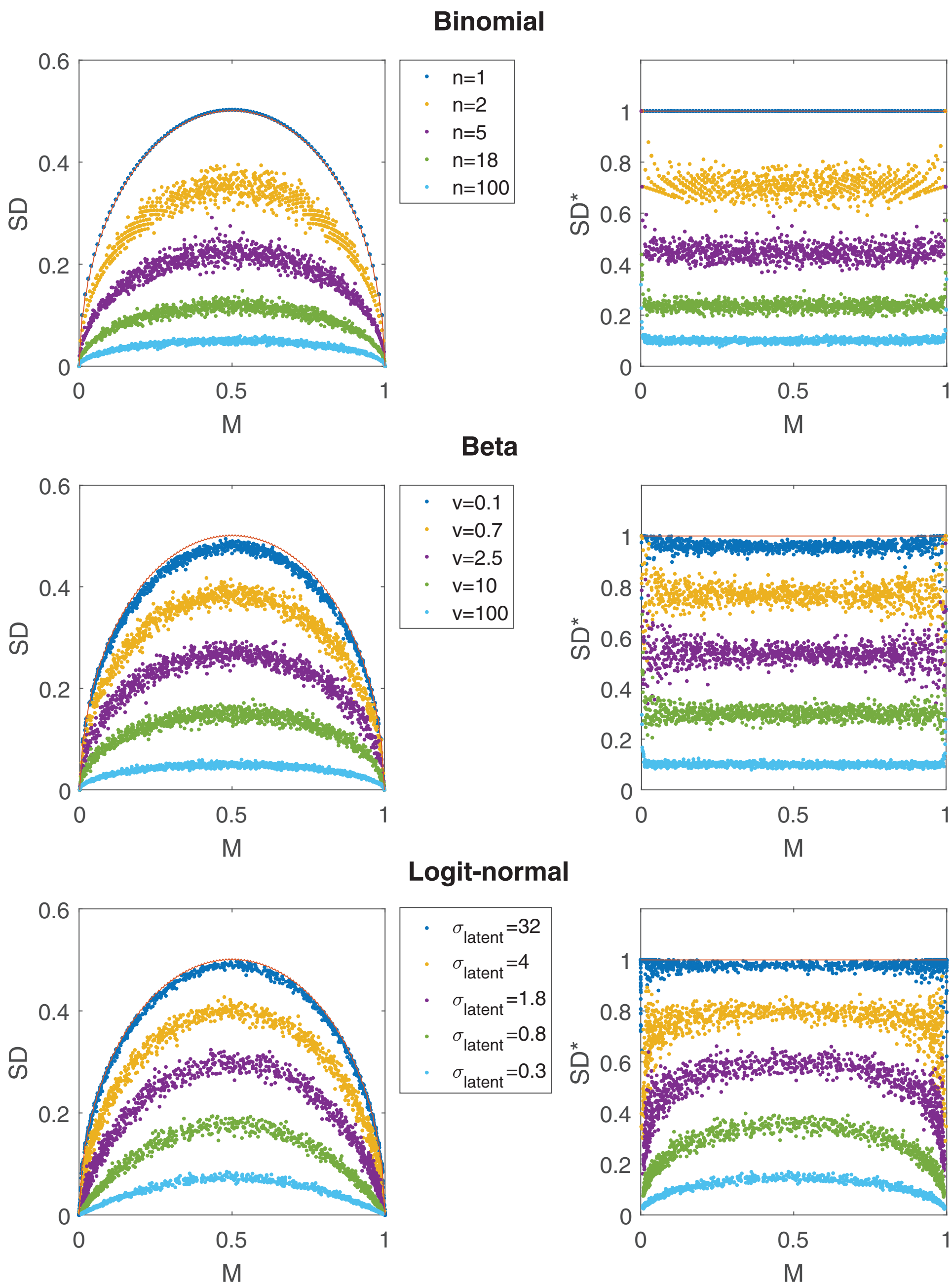

Figure 2. Simulation of different distributions with random means and with number of measures per time series $N_i = 100$. In the left panels we show the mean against the standard deviation and on the right panels we show the mean against the relative standard deviation. Top panels: Samples from binomial distributions as shown in Equation 7. Points with the same color come from a distribution with the same number of trials *n*. Note that we display the proportions (number of successes divided by number of trials). The top distributions in blue are the Bernoulli distributions, where the maximum variability is always exactly achieved (i.e., a binomial with a single trial). Middle panels: beta distributions with different concentrations or precisions ν (same color refers to distributions with same concentration). Bottom panels: logit normal distributions (points with same color come from distributions with same latent standard deviation σ_{latent}).

0.77 while the correlation between σ_{latent} and the *SD* is only 0.50. As the correlation between the *SD* and the SD^* is again 0.77 one could state that the use of the σ_{latent} of the logit normal distribution as a more extreme correction as the SD^*.

In conclusion, our analysis shows that for several bounded distributions, the mean and the standard deviation are inherently mathematically related while the mean and the relative standard deviation are not. This holds for the Bernoulli, binomial, beta, and the logit normal distribution. For the first three, the dependence is completely removed when using the relative standard deviation. If the underlying distribution is logit normal, the relative standard deviation will take out a substantial part of the mathematical dependency, but not all, and hence some form of dependence remains.

Simulation Study

Before we apply the relative variability measure to real data sets we want to test how it behaves in cases where we know how an individual difference or trait variable (e.g., depression, denoted generically as D in the simulation study) is related to the mean and variability of a time series in a sample of persons.

First, we created several time series data sets. Each data set included $K = 50$ (i.e., number of participants) series of length $N_i = 50$, assuming a response scale that runs from zero to 100. Each data set was created with one of the options shown in Table 1 and Figure 3. For each data set, we sampled 50 means and variabilities from different ranges, depending on the option. To investigate the effect of data sets with means close to the bounds we simulated data sets both with means close to the bounds as well as with means far from the borders zero and 100. For most options we chose independent mean and variabilities, but for obvious reasons previously discussed, this is not possible when the standard deviation SD is used. Using the sampled means and variabilities we then simulated data sets with the beta distributions or the inverse logit of latent normal distributions. All data were then rounded to the nearest integer to imitate widely used ESM scales.

Second, as shown in Table 1, we then created an outcome variable D that is linearly dependent on the mean and the variability. For all options in Table 1 (to test for the Type I error), we also created data sets where the dependent variable was independent of the variability measure, after controlling for the mean. For each option we created 1,000 data sets with and without an effect.

Finally, we tested if we could find an effect of the variability using the following six analysis models:

$$D_i = \beta_0 + \beta_1 SD_i + \varepsilon_i \tag{8a}$$

$$D_i = \beta_0 + \beta_1 SD_i + \beta_2 M_i + \varepsilon_i \tag{8b}$$

$$D_i = \beta_0 + \beta_1 SD_i^* + \beta_2 M_i + \varepsilon_i \tag{8c}$$

$$D_i = \beta_0 + \beta_1 SD_i + \beta_2 M_i + \varepsilon_i \text{ (weighted regression)} \tag{8d}$$

$$D_i = \beta_0 + \beta_1 SD_i^* + \beta_2 M_i + \varepsilon_i \text{ (weighted regression)} \tag{8e}$$

$$D_i = \beta_0 + \beta_1 SD_{i,latent} + \beta_2 M_{i,latent} + \varepsilon_i \tag{8f}$$

So this means that simulation Options 1–10 (each with and without an effect of variability) are crossed with the analysis Models a–f. In Equations 8a–8f, M_i, SD_i and SD_i^* are the computed mean, standard deviation, and relative standard deviation for the observed time series of individual i. In addition, $SD_{i,latent}$ and $M_{i,latent}$ are the estimated mean and the standard deviation of the latent logit normal distribution, using the likelihood as described in Equation 3.

We included options where the outcome variable D is simulated using the normal SD the relative SD^* and the latent SD_{latent}, such that Models b, c, and f are all correct (and wrong) in a subset of the simulations. Model b is the model often assumed in contemporary data analysis (Koval, Pe, et al., 2013), Model 8c is the model proposed in this paper and Model 8f uses the $SD_{i,latent}$, an alternative solution for the mean variability dependency problem. We did not include an option where the outcome variable D is independent of the mean M as this seemed unrealistically for most psychological contexts, therefore Model 8a is never correct. Models 8d and 8e where included to investigate the weighted approach we recommend when data with means near the bounds cannot be trusted or one does not want to fully commit to one of the proposed models, which differ the most from each other at the bounds.

To compare the models, we calculated in what proportion we find a significant effect (i.e., using a significant threshold of 0.05) in the correct direction.

Table 2 shows the results. Generally, in most cases the correct model leads to the highest power, or at least the power of the correct model is close to the best. In addition, if there is no effect, the true model has a Type I error rate equal to the nominal 5%.

When the mean M is simulated from a range away far from the borders zero and 100, the models do not differ much from each

Table 1
The Different Options Discussed in the Simulation Study

Option	Mean	Variability	x_i	Simulation model
1	$M_i \sim U(25, 100)$	$SD_i^* \sim U(0.4, 0.8)$	Beta	$D_i = M_i(+200SD_i^*) + 50\varepsilon_i$
2	$M_i \sim U(25, 75)$	$SD_i^* \sim U(0, 0.8)$	Beta	$D_i = M_i(+100SD_i^*) + 50\varepsilon_i$
3	$M_i \sim U(75, 100)$	$SD_i^* \sim U(0.4, 0.8)$	Beta	$D_i = M_i(+200SD_i^*) + 50\varepsilon_i$
4	$M_i \sim U(25, 100)$	$SD_i \sim U\left(\frac{50 - \frac{M_i}{2}}{2}, 50 - \frac{M_i}{2}\right)$	Beta	$D_i = M_i(+2SD_i) + 50\varepsilon_i$
5	$M_i \sim U(40, 60)$	$SD_i \sim U(20, 40)$	Beta	$D_i = M_i(+4SD_i) + 50\varepsilon_i$
6	$M_i \sim U(75, 100)$	$SD_i \sim U(0, 100 - M_i)$	Beta	$D_i = M_i(+2SD_i) + 50\varepsilon_i$
7	$M_i \sim U(25, 100)$	$SD_i^* \sim U(0.4, 0.8)$, $SD_i = SD_i^* \max(SD_i \mid M_i)$	Beta	$D_i = M_i(+2SD_i) + 50\varepsilon_i$
8	$M_{i,latent} \sim U(-4, 8)$	$SD_{i,latent} \sim U(2, 6)$	Logit normal	$D_i = M_{i,latent}(+10SD_{i,latent}) + 50\varepsilon_i$
9	$M_{i,latent} \sim U(-5, 5)$	$SD_{i,latent} \sim U(0, 6)$	Logit normal	$D_i = M_{i,latent}(+10SD_{i,latent}) + 50\varepsilon_i$
10	$M_{i,latent} \sim U(2, 4)$	$SD_{i,latent} \sim U(0, 2)$	Logit normal	$D_i = M_{i,latent}(+30SD_{i,latent}) + 50\varepsilon_i$

Note. Columns 2 (mean) and 3 (variability) show the distributions (and corresponding ranges) from which the mean and variabilities are drawn. These distributions and ranges are visualized in Figure 3. Column 4 (x_i) shows the distributions from which the time series are simulated, using the previously defined mean and variability. Note that all x_i are rounded to the nearest integer. In the last column (simulation model) it is shown how the outcome variable D is created (the parentheses indicate that data are simulated with and without an effect of variability).

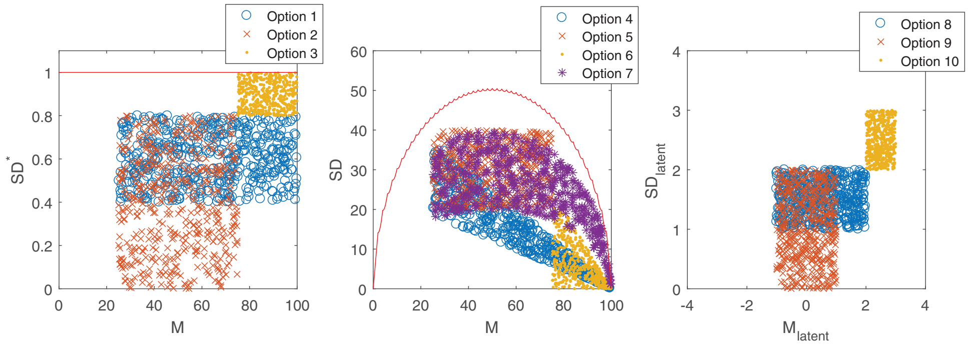

Figure 3. The areas used to simulate means and variabilities from, as described in Table 1. For this figure we simulated 500 means and variabilities for each option. Throughout all options, three kind of true models were used. One using the SD^{*} for which the ranges are shown in the left panel, one using SD for which the ranges are shown in the middle panel and one using the SD_{latent} for which the ranges are shown in the right panel. For each model we included one mean area that concentrates around the middle where $M = 50$, far from the upper bound 100 and lower bound 0, one area which covers the middle (50) and ranges up to the upper bound and one which concentrates in the area just under the upper bound.

other. The relative change from the normal standard deviation to the relative standard deviation is small. This is the case in Option 2, 5 and 9 and leads to only a small difference in power between the possible models.

For Options 1 to 3, the true model, which also performs best, is model c, a linear regression including the mean M and the SD^{*}.

For Options 4 to 7, Model b is the true model, and this is a linear regression of the mean and the normal standard deviation. However, the standard deviation only clearly outperforms the relative standard deviation SD^{*} when we simulate a SD related to the maximum possible standard deviation as in Option 7. Although Options 4 and 6 also include means M near the bounds, Model b does not outperform Model c. If the standard deviation decreases linearly to zero, the SD does not differ much from the SD^{*}; the average correlation between the SD^{*} and the SD is respectively 0.97 and 0.88 for Options 4 and 6. In

Table 2
The Results From the Simulation Study That Follows From Crossing the 10 Simulations Models (With and Without an Effect of Variability)

Option	Effect	a: SD	b: SD and M	c: SD^{*} and M	d: Weighted SD and M	e: Weighted SD^{*} and M	f: SD_{latent} and M_{latent}
1	Yes	.05	.55	**.77**	.66	.76	.58
	No	.35	.04	.05	.05	.04	.09
2	Yes	.85	.88	**.89**	.88	.88	.82
	No	.04	.04	.04	.04	.04	.04
3	Yes	.09	.08	**.21**	.13	.19	.06
	No	.12	.06	.06	.05	.04	.05
4	Yes	.07	.10	.11	**.11**	**.11**	.10
	No	.76	.05	.04	.06	.06	.07
5	Yes	.80	**.82**	.81	**.82**	**.82**	.71
	No	.04	.04	.04	.04	.04	.05
6	Yes	.13	.18	.16	**.21**	**.21**	.16
	No	.08	.05	.05	.07	.06	.04
7	Yes	.08	**.34**	.20	.26	.24	.13
	No	.33	.05	.05	.06	.05	.09
8	Yes	.27	.33	.38	.33	.38	**.40**
	No	.07	.06	.06	.06	.06	.06
9	Yes	**.33**	.32	.32	.32	.32	**.33**
	No	.06	.06	.06	.06	.06	.05
10	Yes	.50	.40	.41	.40	.41	**.74**
	No	.06	.05	.05	.05	.05	.05

Note. Column 1 shows the option that was used to simulate the data as described in Table 1. The second column shows weather there was an effect of the variability or not. The other columns show the proportion of significant results (at the 5% level) for the regression coefficient of variability for the different models a–f from Equation 8. If there was an effect, only the proportion of data sets that found an effect in the right direction are shown. If there was no effect, the proportion of data sets that found any effect is calculated. If there was an effect, the highest power is shown in bold. The shaded cells refer to the situation where the analysis model is the true model.

Option 7 on the other hand, the correlation between the SD^* and the SD is only 0.59. Such that when one assumes that there is an effect of the SD it is more difficult to find it using the SD^*. Note that only in situations like Option 5, with no observations near the bounds, we can simulate M and SD independently. In the other options, the SD is somehow related to the mean M such that the effect of the SD on the outcome variable D could also be interpreted as an indirect effect of the mean. Model b might be able to handle the linear dependency between the mean M and the SD in Options 4 and 6 but not the nonlinear dependency of Model 7. In this sense, the increased power of Model b for Option 7 may partly be seen as incorrectly modeling an effect of the mean M as variability.

In Options 8 to 10, the true model is model f, a linear regression including M_{latent} and SD_{latent}. As we saw in the previous section that the SD_{latent} can be seen as a more extreme version of the SD^* it is therefore logical that models including SD^* slightly outperform models including SD. In Option 10, with only means near the bounds, SD_{latent} is most different from the SD and the SD^*. The inflation of the SD_{latent} compared with the normal SD is even more extreme as the SD^*. Near the bounds, this inflation is highest, so here it differs most from the normal SD and the relative SD^*. If the outcome variable D is assumed to be related to this inflated variability measure, both the SD as the relative SD^* will underperform as shown in Table 1.

Using Model 8a, where the outcome variable is assumed to be independent of the mean M, is clearly a bad idea if the outcome variable is in fact dependent on the mean. Neglecting the mean leads to inflated Type I errors and decreased power.

In general, for the correct model, a weighted analysis leads to a lower power, but for the wrong model a weighted analysis may even lead to a higher power, such as in Options 1 to 3 and 7. The observations with means near the border are weighted less, and these are exactly the observations which deviate the most between the models. Despite not being shown in Table 2, it should be noted that that, without the use of the recalculated sample size as in Equation 6, also the weighted regression lead to increased Type I errors (i.e., over 0.1).

Besides Model a, Model f leads to the highest Type I errors. For high variabilities the estimates of SD_{latent} have a high variance leading to heteroscedasticity, which in turn may lead to an increased Type I error.

Empirical Results

We will now apply our measure to five real-life data sets. First, using three so-called big data examples, we will show the omnipresence of the mean-variability dependency and how this dependency is avoided by using the relative variability. Second, using two smaller data sets, we will show how this dependency can influence conclusions if it is not treated properly. For illustrative purposes, in each application we will focus on the SD as a variability measure.

The Relative Variability Index in Big Data

We will first examine the relation between the mean M and the standard deviation SD in three large data sets. We will show how the SD is clearly related to the M while the SD^* is not. To assess the relation between two variables, four methods will be used. First, the Pearson correlation ρ_{lin} is used to investigate the linear relation between the mean and, respectively, the SD and the SD^*. If the averages of the individuals are spread over the whole range, as in Figure 2, we do not expect any linear correlations. Sometimes however, most individuals have a mean closer to one of the bounds, which may in fact lead to high Pearson correlations. Second, because the Pearson correlation is only able to pick up the linear dependence between the mean and the standard deviation, we also use the distance correlation ρ_{dist} (Székely, Rizzo, & Bakirov, 2007) to examine their nonlinear dependence. The distance correlation between two variables is a measure of both statistical linear and nonlinear dependence and ranges from zero (both variables are independent) to one. Third, as we mainly expect that the relation between the mean and the variability measure is the result of the maximum possible variability $\max(SD|M)$, we also calculate the Pearson correlation between this $\max(SD|M)$ and both variability measures, the SD and the SD^*. Moreover, using big data sets gives us a fourth and more intuitive way to assess the relation between the mean and the variability measure: As each data consists out of a large number of data points, dependence can be simply evaluated by inspecting graphical representation of the data.

In a first big data example, we analyze data which was collected through an online dating site (http://libimseti.cz/, Brozovsky & Petricek, 2007). The data involve a total of 135,359 participants, each rating at least 20 other participants as potential partners on their attractiveness. Each rating is bounded between 1 and 10, where 10 is the best possible attractiveness rating. Participants who provided constant ratings were excluded. Together, all participants made 17,359,346 online ratings. The intraindividual mean M_i, standard deviation SD_i and relative standard deviation SD_i^* were computed for each participant. The relation between the mean and both variability measures is shown in Figure 4. Because of the large amount of data, the shape of the maximum possible standard deviation as a function of the mean can be traced out fairly well when the SD is used. On the other hand, there seems to be no relation between the M and the SD^*. This can also be concluded using the calculations of the Pearson correlation ρ_{lin} the distance correlation ρ_{dist} and the correlation between the standard deviation and its maximum (i.e., ρ_{exp}) as shown in Table 3. The SD^* is clearly much more orthogonal to the mean M.

In a second big data example we study a data set collected using the free smartphone application "58 seconds" that monitored emotions in daily live (Trampe, Quoidbach, & Taquet, 2015). Using the application, participants answered whether they experienced a certain emotion or not. We will focus on the number of different positive emotions (alertness, amusement, awe, gratitude, hope, joy, love, pride, and satisfaction) participants felt at each specific time point. We included 1,566 participants who used the application at least 10 times, resulting in 33,862 completed questionnaires. Again, the intraindividual mean M_i, standard deviation SD_i, and relative standard deviation SD_i^* were computed for the number of positive emotions reported across measurement occasions for each subject. The relation between the mean and both variability measures is visualized in Figure 5. While it is clearly visible that there is a relation between the M and the SD, the M and the SD^* are independent from each other. This is also reflected by the ρ_{lin} the ρ_{dist} and the ρ_{exp} as shown in Table 3.

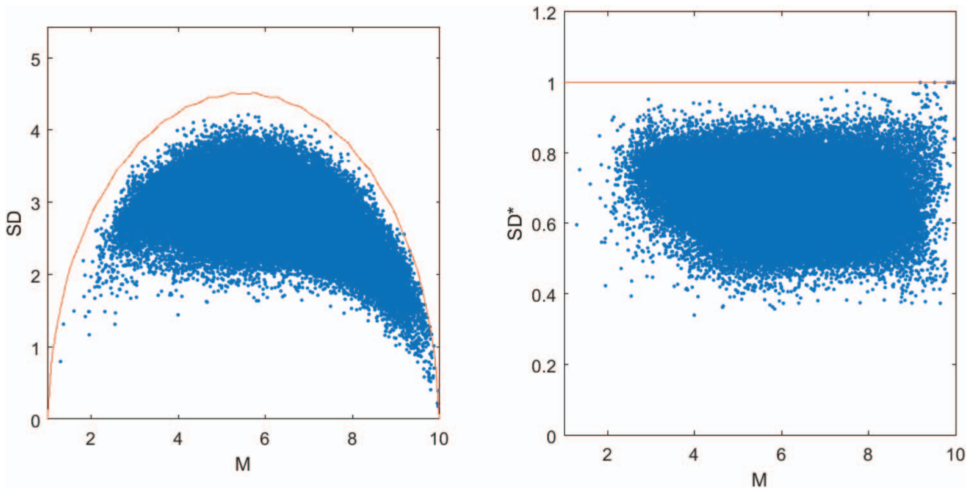

Figure 4. Online dating data set (Brozovsky & Petricek, 2007). Variability of the user ratings versus the average rating. On the left panel the normal standard deviation is used and on the right panel the relative standard deviation is used. The red line is the maximum possible variability, max($SD|M$) or max($SD^*|M$), given the mean for the median number of ratings. In the left panel one can clearly see the inverse U-shape, following max($SD|M$). See the online article for the color version of this figure.

In a third big data example we show that this effect is not limited to intraindividual variability measures, using data reflecting movie ratings collected through the web site Movie Lens (https://movielens.org/, Movie Lens, n.d.). We examined the data of 2,019 movies which each had at least 100 ratings. In total 942,225 ratings where included, ranging from 1 to 5 stars. For each movie we computed the average rating M_i, and the variability in the ratings using the standard deviation SD_i and relative standard deviation SD_i^*. In this case of course, a data point does not refer to a person but to a movie. The relation between the mean and the variability measures is shown in Figure 6. This relation is quantified using the ρ_{lin} the ρ_{dist} and the ρ_{exp} in Table 3. For each of the three statistics, the conclusion is the same. The *M* is much less related to the SD^* as to the *SD*.

We have now shown in three different large data sets across three research contexts, we can expect a relation between the mean *M* and the standard deviation *SD* with bounded scales. This implies that any finding concerning a variable's variability may be driven by the mean, possibly leading to erroneous conclusions. In the following two examples, we will illustrate how the use of the relative variability can indeed lead to different results.

Table 3
The Relation Between the Mean M and the Variability Measures SD and SD^* Quantified Using the ρ_{lin} (Pearson Correlation Between M and SD or SD^*), the ρ_{dist} (the Distance Correlation Between M and SD or SD^*) and the ρ_{exp} (the Pearson Correlation Between SD or SD^* and the max($SD|M$)

	ρ_{lin}		ρ_{dist}		ρ_{exp}	
Data set	*SD*	SD^*	*SD*	SD^*	*SD*	SD^*
Online dating	−.52	−.25	.51	.29	.57	−.012
Smartphone application	.66	.08	.69	.20	.72	−.07
Movie rating data set	−.56	−.22	.57	.25	.55	−.16

Note. This relation is examined for the online dating data set (Brozovsky & Petricek, 2007), the smartphone application data set (Trampe et al., 2015) and the movie rating data set (Movie Lens). For each data set, the *M* is much less related to the SD^* as to the *SD*. As the online dating data set was too big to calculate the distance correlation, we estimated ρ_{dist} using 100 random subsamples of 2,000 subjects.

Application 1: Variability in Thoughts and Feelings

For this application, we analyzed an experience sampling (ESM) data set, in which 95 participants were asked 10 times a day for 7 consecutive days about their momentary feelings and thoughts (Koval, Pe et al., 2013; Pe, Koval, & Kuppens, 2013). In total, 20 variables were measured using sliders (such as anger, depression, happiness, self-esteem, and several appraisals of the situation), each bounded between 1 and 100.

In these data, we are interested in the relation between within-person variability and an individual difference variable (for a review of research related to such questions, see Houben et al., 2015) such as, for instance, depressive symptom severity *D* (as measured with the Center of Epidemiological Studies Depression Scale; CES-D; Lewinsohn, Seeley, Roberts, & Allen, 1997). In general, previous research has found that not only average levels of psychological functioning but also its variability over time is related to well-being (Houben et al., 2015; Kernis & Goldman, 2003). As previously explained, participants with a mean value equal to one of the bounds will be excluded (typically one or two participants depending on the specific variable).

The relation between the intraindividual standard deviation and a third variable is traditionally studied by linear regression from Equation 8a. However, if the mean is not taken into account, any effect of the standard deviation may just be an indirect effect of the mean. Therefore, sometimes, also Equation 8b is used. Still, as we explained above, it is not recommended to use this linear correction if the mean and the standard deviation are related to each other. To see how the

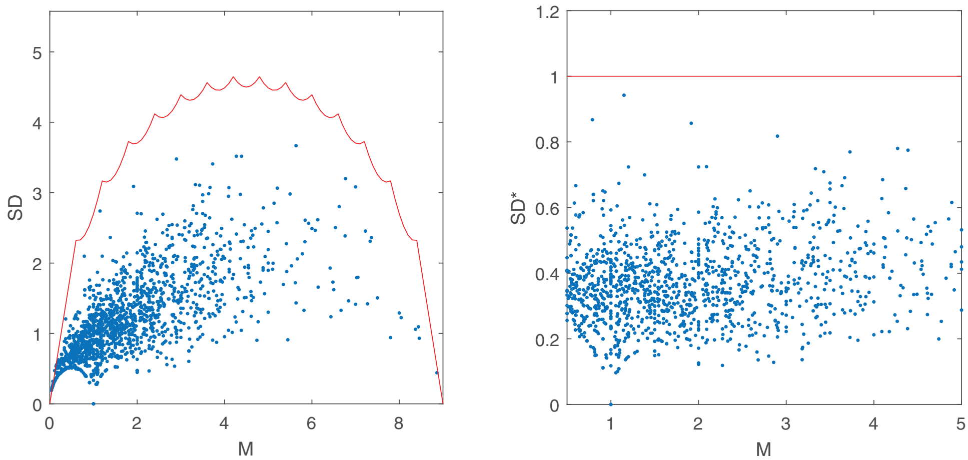

Figure 5. Smartphone application data set (Trampe et al., 2015). Variability of the number of positive emotions versus the average number of positive emotions. On the left panel the normal standard deviation is used and on the right panel the relative standard deviation is used. The red line is the maximum possible variability, max($SD|M$) or max($SD^*|M$), given the mean for the median number of completed questionnaires. See the online article for the color version of this figure.

standard deviation SD is confounded by the mean M for each variable, we again used the three correlation indices introduced above: ρ_{lin} the ρ_{dist} and the ρ_{exp}.

For the linear dependence, we found that the linear correlations ρ_{lin} range from -0.42 to 0.82 across the 20 variables. This outcome is striking for several reasons. First, the correlation between the mean and standard deviation is strongly variable-dependent, with correlations ranging from very negative to extremely positive. Second, due to the presence of large correlations for some ESM variables, using a linear regression approach to correct for the confound can result in multicollinearity. For example, a correlation of 0.82 leads to a variance inflation factor of 3, which means that the variances of regression weights β_1 and β_2 are inflated with a factor of three due to the multicollinearity. This can result in a serious decrease of power, skewing conclusions from such analysis.

We also examined the nonlinear relation between the mean and the variability using the ρ_{dist} and the ρ_{exp}. For this data set, we found that the distance correlation ρ_{dist} between the M and the SD ranged from 0.15 to 0.81 across the 20 variables. The ρ_{exp} ranged from 0.08 to 0.83. Any nonlinear dependencies between the M and the SD are completely ignored in Equation 8b.

Coefficient β_1 of Equation 8b is commonly understood as the expected change in CESD score if SD increases with one unit

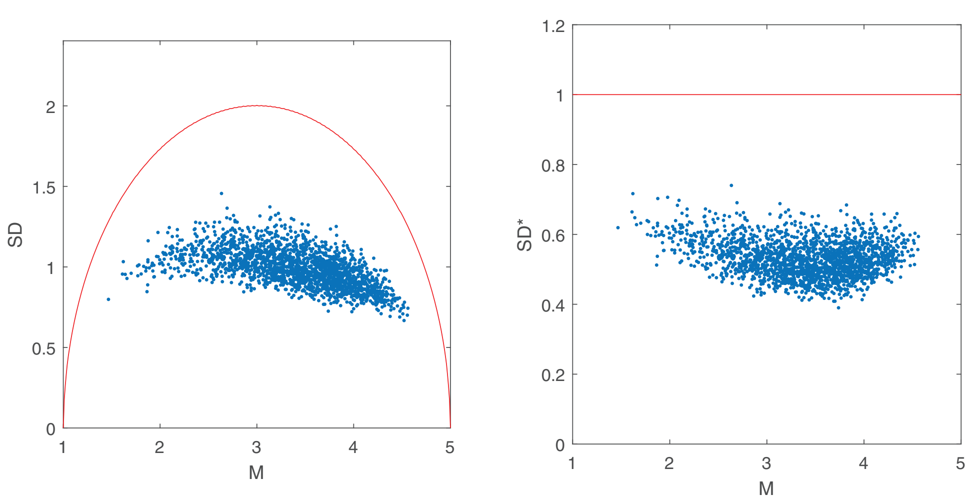

Figure 6. Movie Lens data (Movie Lens, n.d.). Variability of the movie ratings versus the average movie ratings. On the left panel the normal standard deviation is used and on the right panel the relative standard deviation is used. The red line is the maximum possible variability, max($SD|M$) or max($SD^*|M$), given the mean for the median number of ratings per movie. In the left panel one can clearly see the inverse U-shape, following max($SD|M$). See the online article for the color version of this figure.

while M is held constant. Due to the strong linear and nonlinear dependencies between the mean M and the variability SD, this is a contrived interpretation. It makes no sense to keep the M constant while SD increases.

In a next step, the relative standard deviation SD^* was calculated. For the linear dependence (using ρ_{lin}) between the SD^* and the M, values were found that range from −0.21 to 0.28. As can be seen in the left panel of Figure 7, for almost every variable, the absolute value of the correlation between the variability and the mean M decreased when we used the SD^* instead of the SD. Across all variables, the absolute value of the correlation decreased on average from 0.49 to 0.12 (using a paired t-test, this result is significant: $t(19) = 6.605$, $p < 10^{-5}$).

Also the ρ_{dist} and the ρ_{exp} between the M and the SD^* were much lower as compared with the distance correlation between the M and the SD. Across all the variables, the ρ_{dist} decreased on average from 0.51 to 0.22 (using a paired t test, this result is significant: $t(19) = 6.683$, $p < 10^{-5}$). The absolute value of the ρ_{exp} decreased on average from 0.56 to 0.10 (using a paired t test, this result is significant: $t(19) = 8.673$, $p < 10^{-7}$).

Note that these findings are not limited to the transition of the standard deviation to the relative standard deviation. For other variability measures, we find the similar results. For both the RMSSD as the IQR, the ρ_{lin}, the ρ_{dist} and the ρ_{exp} decrease significantly across all the variables if their relative variability variant is used instead (see Table 4).

The relative standard deviation and other relative variability measures are clearly much less confounded (linearly and nonlinearly) with the mean compared to the traditional variability measures. We propose therefore to study the relation of the variability of the ESM variables with depressive symptom severity using Equation 8c. As the relative standard deviation is much less confounded by the mean as the standard deviation, we can now say that any results derived from this equation are less influence by the mean as the results derived from Equation 8b. It is however notable that the use of Equation 8c may lead to straight out different conclusions.

The results presented thus far jointly concern the 20 items. To illustrate the results in a more vivid way, let us select one specific ESM item for a more detailed analysis: self-esteem. Both the mean level of self-esteem as well as the degree of variability over time has been considered important for well-being and depression (Franck et al., 2016; Hayes, Harris, & Carver, 2004; Kernis & Goldman, 2003; Roberts & Gotlib, 1997; Wagner, Lüdtke, & Trautwein, 2015), although the exact role remains unclear (Sowislo, Orth, & Meier, 2014). The Pearson correlation between the M and the SD of self-esteem is −0.27. If the relative standard deviation is used instead as in Equation 8b, the Pearson correlation with the mean changes to −0.08. Similarly, the distance correlation between the predictors decreases from 0.4 to 0.25 and the ρ_{exp} changes from 0.49 to −0.12 if the SD is replaced by the SD^*.

When Equation 8a is used for relating depression to the variability of self-esteem it seems to be the case that variability of self-esteem is related to depressive symptoms, $\beta_1 = 0.029$, $SE = 0.011$, $t(93) = 2.745$, $p = 0.007$. However, by using Equation 8b (linearly controlling for the mean), we find that this effect seems to be an indirect effect of the mean. We find that the explained variance of depressive symptoms rises by only 1%, as compared with a regression with only the mean as predictor, resulting in a total variance explained of 37%. The contribution of the standard deviation in explaining depressive symptoms above and beyond the mean is nonsignificant ($\beta_1 = 0.013$, $SE = 0.009$, $t(92) = 1.416$, $p = 0.160$). However, when the relative standard deviation is used (Equation 8c), the explained variance of depressive symptoms rises with a somewhat larger 4% (compared with 1%), resulting in a total variance explained of 40%. The relative standard deviation is thus able to explain 3% more of the variance of depressive symptoms. The effect of SD^* on depression is now again suggestive of evidence against the null hypothesis of no contribution ($\beta_1 = 0.097$, $SE = 0.438$, $t(92) = 2.503$, $p = 0.014$), providing some evidence for the variability of self-esteem as a predictor of levels of depressive symptoms. This result is not just a side effect of the possible inflation or blow-up of the relative standard deviation at the bounds. If we use a weighted linear

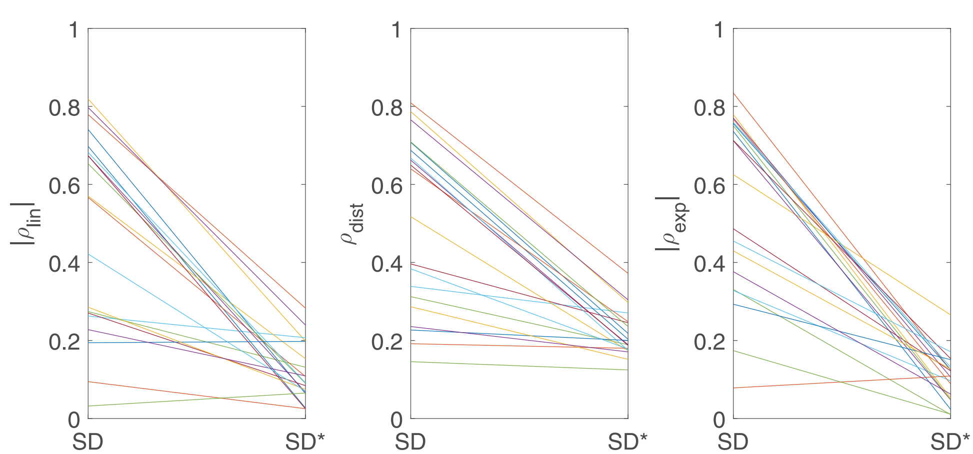

Figure 7. For 26 variables the relation between first the M and SD and second the M and the SD^* is shown. From left to right first the absolute value of the ρ_{lin} then the ρ_{dist} and last the absolute value of the ρ_{exp} is shown. The SD^* is clearly much more independent from the mean compared to SD. See the online article for the color version of this figure.

Table 4
The Relation Between the Mean and Several Variability Measures of 26 Emotion Items From Application 1 (See Text for More Information)

	$\|\rho_{lin}\|$			ρ_{dist}			ρ_{exp}		
Index	Min	Max	Average	Min	Max	Average	Min	Max	Average
SD	.03	.82	.49	.15	.81	.51	.08	.83	.56
*SD**	.03	.28	.12	.12	.37	.22	.01	.27	.10
RMSSD	.00	.72	.41	.16	.72	.45	.03	.78	.52
*RMSSD**	.00	.32	.15	.15	.34	.24	.00	.40	.24
IQR	.21	.91	.56	.31	.90	.61	.12	.82	.56
*IQR**	.15	.57	.34	.26	.54	.38	.01	.51	.26

Note. Quantified using the dependencies $|\rho_{lin}|$, the ρ_{dist} and the ρ_{exp}. We compute for each variability measure the minimum, the maximum and the average of the dependencies over 26 variables. If relative variability measures are used instead of classical variability measures, the $|\rho_{lin}|$, the ρ_{dist} and the ρ_{exp} decrease.

regression with the weights of Equation 5 we find similar results ($\beta_1 = 1.017$, $SE = 0.476$, $t(86.98) = 2.1366$, $p = 0.035$). Here the noninteger degrees of freedom are computed using Equation 6.

At last, to compare our method with a more involved method, we discuss the results using an underlying logit normal distribution. Applying the coarsened approach of Lesaffre et al. (2007), we estimate the latent mean M_{latent} and the latent standard deviation SD_{latent} for each participant. Already in the simulations in Figure 2 we showed that the logit normal distribution leads to an even larger correction near the boundaries than the relative variability measure. This translates itself in the self-esteem data as three clear outliers which are shown in Figure 8. When these outliers are included the Pearson correlation between M_{latent} and SD_{latent} is 0.77. When they are excluded, as we propose to do in the following analysis, the Pearson correlation reduces to a more realistic −0.03. The relation of SD_{latent} with the relation of depression can again be studied using the regression of Equation 8f. Also here, the effect of SD_{latent} on depression is suggestive of evidence against the null hypothesis of no contribution ($\beta_1 = 0.311$, $SE = 0.125$, $t(89) = 2.494$, $p = 0.014$). This conclusion does not change if the outliers are included.

Ignoring the mean, it seems to be clear that the variability of self-esteem (as measured by the traditional standard deviation) contributed significantly to explaining levels of depressive symptoms. However, over and above the mean, there is no evidence that variability of self-esteem contributes in explaining levels of depressive symptoms. When using the relative standard deviation, more variance in depressive symptoms was explained. A similar result can be found by assuming a latent logit normal distribution. The latter approach is however more complicated, overcorrects and inflates variabilities of several participants and leads to different values of the mean *M* as are normally used which makes it backward incompatible.

Application 2: Emotional Variability in Borderline Personality Disorder

In the next application we will again examine ESM data, in which 34 inpatients diagnosed with borderline personality disorder (BPD) and 30 matched healthy participants were prompted 10 times a day, during 8 days to complete a questionnaire about the emotions and thoughts they were currently experiencing (Houben et al., 2016). In total, 18 momentary variables were measured using sliders (such as anger, depression, excitement, and disappointment), each bounded between 0 and 100.

Emotion dysregulation is commonly seen as a core symptom of BPD (Lieb, Zanarini, Schmahl, Linehan, & Bohus, 2004; Linehan, 1993). In fact, emotional instability is an important diagnostic

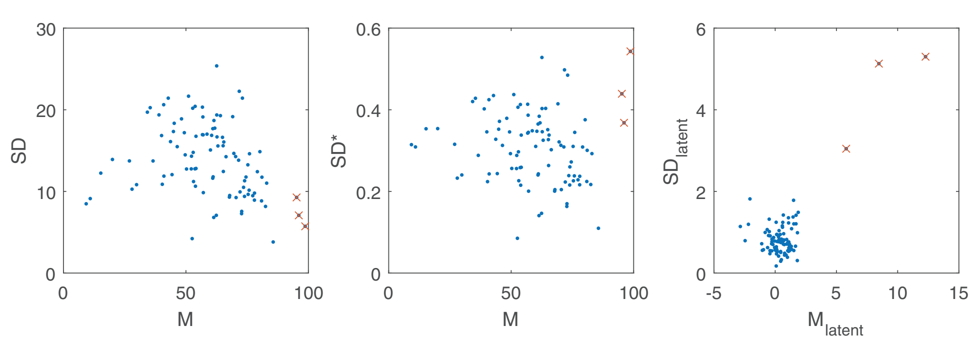

Figure 8. The relation between the mean and standard deviation for self-esteem. From left to right we first show the mean and the normal standard deviation, then the mean and the relative standard deviation, and last the latent mean and the latent standard deviation of the logit normal distribution. One can visually distinguish three outliers in the last plot. These points are shown in all plots using red crosses. See the online article for the color version of this figure.

criterion for BPD (American Psychiatric Association, 2013). Indeed, in a recent meta-analysis Houben et al. (2015) found that BPD is significantly characterized by high levels of emotional variability, particularly of negative emotions. Consequently, here we will examine to what extent individuals diagnosed with BPD display higher levels of negative emotion variability compared with healthy controls, both for a composite negative affect scale (calculated as the average of the ESM variables anger, depressive, anxious, stressed, disappointed in the self, and disappointed in another), and for one specific negative item (i.e., depressed). We hypothesize that borderline patients have a higher variability, more specifically a higher standard deviation, in depressive feelings and negative affect.

To study the relation between emotion variability and borderline personality disorder we make use of the logistic regression. Similarly, as in the previous analysis, in a first step, we do not take the mean into account:

$$\text{odds}(\text{borderline}_i) = e^{(\beta_0+\beta_1 SD_i)}, \quad (9)$$

(where Pr(borderline$_i$) is the probability that person i has BPD) and SD_i is the standard deviation of depressed feelings or negative affect for subject i. Using Equation 9, for both the depressed feelings ($\beta_1 = 0.172$, $SE = 0.045$, $t(52) = 3.061$, $p = 0.0001$) and negative affect ($\beta_1 = 0.342$, $SE = 0.089$, $t(58) = 3.061$, $p = 0.0001$) we found a significant difference between the two groups in terms of emotional variability.

It is however possible that this effect is mainly driven by the mean. There is indeed a correlation ρ_{lin} of about 0.65 between the mean and the variability of both the depressed feeling and the negative affect (see Table 5). In such a case, the least one should do is linearly correct for the mean, using equation:

$$\text{odds}(\text{borderline}_i) = e^{(\beta_0+\beta_1 SD_i+\beta_2 M_i)}.$$

Using this logistic regression we find less explicit results. For both depressed feelings ($\beta_1 = 0.108$, $SE = 0.054$, $t(51) = 1.995$, $p = 0.046$) and negative affect ($\beta_1 = 0.220$, $SE = 0.116$, $t(57) = 1.903$, $p = 0.057$), we find only suggestive evidence against the null hypothesis of no contribution for the standard deviation (beyond the mean).

However, also for a logistic regression, for the same reasons as explained for the linear regression, highly dependent predictors should be avoided. For example, the expected correlation ρ_{exp} between the mean and the standard deviation of the feeling depressed is 0.84. This means that $\rho^2_{exp} = 0.84^2 = 0.70$ (or 70%) of the variance of standard deviation can be explained by a function of the mean (i.e., max($SD|M$)). The standard deviation is again confounded by the mean. We therefore propose to use the relative standard deviation which is much less dependent on the mean, as is shown in Table 5. Using equation

$$\text{odds}(\text{borderline}_i) = e^{(\beta_0+\beta_1 SD^*_i+\beta_2 M_i)},$$

there is no more room for discussion. For both depressed feelings ($\beta_1 = -0.573$, $SE = 1.640$, $t(51) = -0.349$, $p = 0.727$) and negative affect ($\beta_1 = 5.8805$, $SE = 4.526$, $t(57) = 1.293$, $p = 0.194$), we find that there is no evidence against the null hypothesis of no contribution of the standard deviation. In addition, for depressed feelings, the regression estimate even switches signs if the relative standard deviation is used instead of the normal standard deviation. Also in this example, the result is robust against the inflation or blow-up effect at the bounds as using a weighted logistic regression leads to the same conclusions for both depressed feelings ($\beta_1 = 3.4033$, $SE = 3.3026$, $t(31.99) = 1.0305$, $p = 0.303$) as negative affect ($\beta_1 = 8.5244$, $SE = 5.97$, $t(39.97) = 1.429$, $p = 0.153$) $\beta_1 = 0.1200$, $SE = 0.177$, $t(49) = 0.676$, $p = 0.499$) $\beta_1 = 0.0830$, $SE = 0.287$, $t(57) = 0.289$, $p = 0.773$).

Disregarding the mean, we would have concluded that a higher standard deviation of depressed feelings and negative affect is clearly a feature of borderline patients. By linearly correcting for the mean, some researchers still might have concluded that the standard deviation of negative emotions is indeed higher for borderline patients. Using the relative standard deviation, however, these data suggest that any difference between groups in standard deviation is just a side effect of the mean. This effect is striking, given that emotional instability is a diagnostic feature of BPD. If this finding is robust and replicated, it may have consequences for the diagnostic criteria of BPD.

Table 5
The Relation Between the Mean M and the Variability Measures SD and SD Quantified Using the ρ_{lin} the ρ_{dist} and the ρ_{exp} for Application 2*

	ρ_{lin}		ρ_{dist}		ρ_{exp}	
ESM variable	*SD*	*SD**	*SD*	*SD**	*SD*	*SD**
Depressed	.62	−.17	.68	.28	.84	−.29
Negative affect	.65	−.12	.69	.21	.79	−.12

Note. This relation is shown here for depressed feelings and negative affect. Results are similar but less pronounced for the other feelings and emotions measured with this ESM data set (Houben et al., 2016).

Discussion

In many areas of psychological research, the causes, consequences, and correlates of levels of variability in feelings, thoughts, and behavior (either within or between individuals) are being investigated. It has been long and well known that when studying questions involving variability, results can be clouded due to confounding of measures of variance with the mean, especially in the case of bounded measurements. Until now, there has been no single solution that can disentangle the effects of variability measures from the effects of the mean in all circumstances. Here, we proposed a novel metric of variability measures, called relative variability. The set of new proposed measures is independent from the mean, and allows researchers to draw precise conclusions about how variability relates to other phenomena.

We derived analytical solutions and provide a software package which allows the fast calculation of the relative variability measure for the standard deviation, the root mean squared successive difference, the interquartile range, and the range.

In five real-life data sets we showcased the dependency between the mean and the standard deviation. For most data sets, more than half of the variance of the *SD* is explained by a function of the mean, clearly demonstrating the need for indices of variability that are free of this confound. Moreover, in two applications involving the relation between variability and third variables, we showed that using the relative variability index may lead to (radically) different conclusions.

Our relative variability index bears a lot of similarity with the coefficient of variation, except that is suited for measures bounded both from below and above (while the coefficient of variation is only used for measures bounded from below).

A number of criticisms may be raised against the relative variability measure. First, the relative variability index may inflate or blow-up variabilities close to the bounds. Of course, this is a direct consequence of the way the index is defined. This property is shared with the coefficient variation. Luckily we know exactly how much this variability index is inflated so we can quantify this extra uncertainty and take it into account in later analyses by an appropriate weighting. Of course, if a researcher would, for some theoretical reasons, be certain that the variability of data near the bounds is correctly measured as close to zero, the inflating relative variability index should not be used. However, such a certainty is a rare privilege for most researchers. In any case we recommend, even if our relative variability is not used, to apply a more conservative weighted analysis to make sure the results are not too influenced by observations close to the bound.

A second criticism of the relative variability measure is that it takes too much for granted that the measurements are continuous and that performing certain operations (e.g., calculating the mean) are meaningful. This has been an assumption we made throughout the article. If there are reasons to doubt the validity of this assumption (e.g., when a 3-point scale is used), then the only thing to do is to use a latent variable model for ordered categories, but with a person-specific (latent) variability. The most straightforward way to fit such models is by turning to the Bayesian framework and by using software such as JAGS (Plummer, 2003).

A crucial assumption of the relative variability index is that each participant uses the given objective bounds or mapping function in her replies. However, if some participants for example refuse to use a part of the scale, their subjective bounds may not be equal to the real objective bounds. This may lead to a loss of the validity of the relative variability. Unfortunately, disentangling different bounds, different mapping functions and differences in variability is not an easy task and one may wonder whether it is possible at all.

In sum, we have proposed the relative variability index as an easy-to-compute and easy-to-understand variability measure without a confound by the mean. In a number of applications, it is shown that the relative variability index serves its purpose.

Code and Data Availability

Our method is implemented in R and MATLAB, the software package and all data and code necessary to replicate the real data examples in this article can be downloaded at http://ppw.kuleuven.be/okp/software/relative_variability/

References

Aitchison, J., & Shen, S. M. (1980). Logistic-Normal Distributions: Some Properties and Uses. *Biometrika, 67,* 261–272. http://dx.doi.org/10.2307/2335470

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM–5*. Washington, DC: American Psychiatric Publishing Incorporated.

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90,* 512–527. http://dx.doi.org/10.1037/0022-3514.90.3.512

Brozovsky, L., & Petricek, V. (2007). *Recommender system for online dating service*. Retrieved from http://www.occamslab.com/petricek/papers/dating/brozovsky07recommender.pdf

Diehl, M., Hooker, K., & Sliwinski, M. J. (2014). *Handbook of intraindividual variability across the life span*. New York, NY: Routledge.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76,* 662–676. http://dx.doi.org/10.1037/0022-3514.76.4.662

Elvira, V., Martino, L., Luengo, D., & Bugallo, M. F. (2017). Improving population Monte Carlo: Alternative weighting and resampling schemes. *Signal Processing, 131,* 77–91. http://dx.doi.org/10.1016/j.sigpro.2016.07.012

Faraway, J. J. (2004). *Linear Models with R*. Boca Raton, FL: Chapman and Hall/CRC.

Fleeson, W. (2004). Moving personality beyond the person-situation debate the challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13,* 83–87. http://dx.doi.org/10.1111/j.0963-7214.2004.00280.x

Fleeson, W., & Law, M. K. (2015). Trait enactments as density distributions: The role of actors, situations, and observers in explaining stability and variability. *Journal of Personality and Social Psychology, 109,* 1090–1104. http://dx.doi.org/10.1037/a0039517

Fleeson, W., & Wilt, J. (2010). The relevance of big five trait content in behavior to subjective authenticity: Do high levels of within-person behavioral variability undermine or enable authenticity achievement? *Journal of Personality, 78,* 1353–1382.

Franck, E., Vanderhasselt, M.-A., Goubert, L., Loeys, T., Temmerman, M., & De Raedt, R. (2016). The role of self-esteem instability in the development of postnatal depression: A prospective study testing a diathesis-stress account. *Journal of Behavior Therapy and Experimental Psychiatry, 50,* 15–22. http://dx.doi.org/10.1016/j.jbtep.2015.04.010

Golay, P., Fagot, D., & Lecerf, T. (2013). Against coefficient of variation for estimation of intraindividual variability with accuracy measures. *Tutorials in Quantitative Methods for Psychology, 9,* 6–14. http://dx.doi.org/10.20982/tqmp.09.1.p006

Hayes, A. M., Harris, M. S., & Carver, C. S. (2004). Predictors of self-esteem variability. *Cognitive Therapy and Research, 28,* 369–385. http://dx.doi.org/10.1023/B:COTR.0000031807.64718.b9

Hedeker, D., Berbaum, M., & Mermelstein, R. (2006). Location-scale models for multilevel ordinal data: Between- and within-subjects variance modeling, *4,* 1–20.

Hedeker, D., Demirtas, H., & Mermelstein, R. J. (2009). A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Statistics and Its Interface, 2,* 391–401. http://dx.doi.org/10.4310/SII.2009.v2.n4.a1

Holland, P. W., & Wang, Y. J. (1986). Regional dependence for continuous bivariate densities. *ETS Research Report Series, 1986,* i-15. http://dx.doi.org/10.1002/j.2330-8516.1986.tb00170.x

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, no pagination specified.*

Houben, M., Vansteelandt, K., Claes, L., Sienaert, P., Berens, A., Sleuwaegen, E., & Kuppens, P. (2016). Emotional switching in borderline personality disorder: A daily life study. *Personality Disorders, 7,* 50–60. http://dx.doi.org/10.1037/per0000126

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13,* 354–375. http://dx.doi.org/10.1037/a0014173

Kalmijn, W., & Veenhoven, R. (2005). Measuring inequality of happiness in nations: In search for proper statistics. *Journal of Happiness Studies, 6,* 357–396. http://dx.doi.org/10.1007/s10902-005-8855-7

Kernis, M. H., & Goldman, B. M. (2003). Stability and variability in self-concept and self-esteem. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 106–127). New York, NY: Guilford Press.

Kong, A. (1992). *A note on importance sampling using standardized weights*. Tech. Rep. No. 348. Department of Statistics, University of Chicago.

Koval, P., Ogrinz, B., Kuppens, P., Van den Bergh, O., Tuerlinckx, F., & Sütterlin, S. (2013). Affective instability in daily life is predicted by resting heart rate variability. *PLoS ONE, 8,* e81536. http://dx.doi.org/10.1371/journal.pone.0081536

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion, 13,* 1132–1141. http://dx.doi.org/10.1037/a0033579

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Cambridge, MA: Academic Press/Elsevier. Retrieved from http://www.indiana.edu/~kruschke/DoingBayesianDataAnalysis/

Kružík, M. (2000). Bauer's maximum principle and hulls of sets. *Calculus of Variations and Partial Differential Equations, 11,* 321–332. http://dx.doi.org/10.1007/s005260000047

Lesaffre, E., Rizopoulos, D., & Tsonaka, R. (2007). The logistic transform for bounded outcome scores. *Biostatistics, 8,* 72–85. http://dx.doi.org/10.1093/biostatistics/kxj034

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., & Allen, N. B. (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and Aging, 12,* 277–287. http://dx.doi.org/10.1037/0882-7974.12.2.277

Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *Lancet, 364,* 453–461. http://dx.doi.org/10.1016/S0140-6736(04)16770-6

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: Guilford Press.

MATLAB. (2016). Version 9.1.0.441655 (R2016b) [Computer software]. Natick, MA: The MathWorks Inc.

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science, 18,* 112–117. http://dx.doi.org/10.1111/j.1467-8721.2009.01619.x

Movie Lens. (n.d.). Retrieved from http://files.grouplens.org/datasets/movielens/ml-1m-README.txt

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *The Journals of Gerontology Series B, Psychological Sciences, and Social Sciences, 59,* 49–55. http://dx.doi.org/10.1093/geronb/59.2.P49

Pe, M. L., Koval, P., & Kuppens, P. (2013). Executive well-being: Updating of positive stimuli in working memory is associated with subjective well-being. *Cognition, 126,* 335–340. http://dx.doi.org/10.1016/j.cognition.2012.10.002

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, MA: Harvard University Press. http://dx.doi.org/10.4159/9780674369542

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. Retrieved from https://www.r-project.org/conferences/DSC-2003/Drafts/Plummer.pdf

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging, 24,* 778–791. http://dx.doi.org/10.1037/a0017915

Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging, 20,* 623–633. http://dx.doi.org/10.1037/0882-7974.20.4.623

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roberts, J. E., & Gotlib, I. H. (1997). Temporal variability in global self-esteem and specific self-evaluation as prospective predictors of emotional distress: Specificity in predictors and outcome. *Journal of Abnormal Psychology, 106,* 521–529. http://dx.doi.org/10.1037/0021-843X.106.4.521

Rocha, A. V., & Cribari-Neto, F. (2008). Beta autoregressive moving average models. *Test, 18,* 529–545. http://dx.doi.org/10.1007/s11749-008-0112-z

Segerstrom, S. C., & Nes, L. S. (2007). Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological Science, 18,* 275–281. http://dx.doi.org/10.1111/j.1467-9280.2007.01888.x

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: CRC Press. http://dx.doi.org/10.1201/9780203489437

Sowislo, J. F., Orth, U., & Meier, L. L. (2014). What constitutes vulnerable self-esteem? Comparing the prospective effects of low, unstable, and contingent self-esteem on depressive symptoms. *Journal of Abnormal Psychology, 123,* 737–753. http://dx.doi.org/10.1037/a0037770

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics, 35,* 2769–2794. http://dx.doi.org/10.1214/009053607000000505

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PLoS ONE, 10,* e0145450. http://dx.doi.org/10.1371/journal.pone.0145450

Wagner, J., Lüdtke, O., & Trautwein, U. (2015). Self-esteem is mostly stable across young adulthood: Evidence from latent STARTS models. *Journal of Personality, 84,* 523–535. http://dx.doi.org/10.1111/jopy.12178

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815–827. http://dx.doi.org/10.1037/0022-3514.81.5.815

Received July 20, 2016
Revision received May 7, 2017
Accepted May 15, 2017 ■