
Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations

Bradley Harding ^a, Christophe Tremblay^a, Denis Cousineau^a

^a School of Psychology, Université d'Ottawa

Abstract ■ Characteristics of a population are often unknown. To estimate such characteristics, random sampling must be used. Sampling is the process by which a subgroup of a population is examined in order to infer the values of the population's true characteristics. Estimates based on samples are approximations of the population's true value; therefore, it is often useful to know the reliability of such estimates. Standard errors are measures of reliability of a given sample's descriptive statistics with respect to the population's true values. This article reviews some widely used descriptive statistics as well as their standard error estimators and their confidence intervals. The statistics discussed are: the arithmetic mean, the median, the geometric mean, the harmonic mean, the variance, the standard deviation, the median absolute deviation, the quantile, the interquartile range, the skewness, as well as the kurtosis. Evaluations using Monte-Carlo simulations show that standard errors estimators, assuming a normally distributed population, are almost always reliable. In addition, as expected, smaller sample sizes lead to less reliable results. The only exception is the estimate of the confidence interval for kurtosis, which shows evidence of unreliability. We therefore propose an alternative measure of confidence interval based on the lognormal distribution. This review provides easy to find information about many descriptive statistics which can be used, for example, to plot error bars or confidence intervals.

Keywords ■ Standard error; estimation; evaluation

 bhard024@uottawa.ca

Introduction

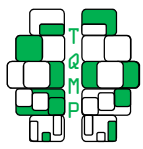
Sampling is the basis for empirical research. Since it is generally impossible to measure all possible subjects from a population, a sample is forcibly used as a substitute. Sampling is the process by which a population's subgroup is used in order to infer the population's actual characteristics (for a review of sampling strategies, see Barreiro & Albandoz, 2001; hereafter we suppose that random sampling is used). But, how representative is a sample with respect to the general population? In other words, to what degree does a sample of n observations truly represent the entire population's scores? Since experimental research does not have the luxury to generate an infinite number of samples to answer this question, estimation methods are used with indicators meant to quantify the reliability of the estimates. One such indicator is the standard error.

The standard error is a statistical indicator of the reliability of a descriptive statistic estimated from a sample. Descriptive statistics include, for example, the mean, the median, the variance, etc. The standard error represents the typical amount of error that can be expected from an estimator. Hence, the smaller the

standard error, the better the estimate is likely to be. Standard errors can be used to delineate an interval likely to contain the population's true characteristic. Such intervals are called confidence intervals (Streiner, 1996).

Sadly, standard errors and standard deviations are often confused in the literature. It was shown in a review article (Olsen, 2003) that 14% of the publications in a medical journal failed to specify their measure of dispersion (whether the standard deviation or the standard error). Whereas the standard deviation is a measure of variability of individual scores within a sample, the standard error is a measure of variability between samples if an infinite number of samples could be drawn from a population (Altman & Bland, 2005; Streiner, 1996).

The standard error of a certain sample statistic is formally defined as the standard deviation of that statistic, assuming that a large number of samples is gathered (Kendall & Stuart, 1977). For example, given the mean score for groups of $n = 30$ students on a statistics exam, the standard error would be given by the standard deviation of all sample means (i.e. means of



groups). Two factors determine the precision of most descriptive statistics. Firstly, the standard errors are proportional to the population's variability; a less variable population leads to more reliable estimates. Consequently, the standard error that is measured should be small as well. Secondly, the measure of standard error is inversely proportional to sample size: larger sample sizes leads to more reliable estimates (to be precise, standard error is inversely proportional to a function of sample size, often \sqrt{n}).

The main objective of this article is to review the standard error estimators associated with widely used descriptive statistics. The present article covers statistics of central tendency, statistics of dispersion, as well as statistics of distribution shape (skewness and kurtosis). These equations give an estimated value to the standard error because they are based on a single sample. In other words, the standard error estimators we present here are estimates of the true standard error; hence a second objective of this article is to compare the sample's standard error to the population's standard error. To estimate the latter, we will rely on Monte-Carlo simulations. For simulations used in the following sections, 10,000 samples are taken from a normal distribution with parameters $\mu = 100$, and $\sigma = 3$ for every sample size varying from $n = 5$ (a very small sample) to $n = 80$ (a fairly large sample) by increments of 5. The sample's descriptive statistics are computed, recorded and another sample is then taken. Finally, the mean of the descriptive statistics simulated is compared to the population's true value. The third objective of this article is to review the confidence interval equations based on standard errors and evaluate them with the above simulations. To do so, we compared the span containing 95% of the 10,000 estimates to the 95% confidence intervals.

The following sections examine the most commonly used descriptive statistics: the central tendency statistics (Section 1), the dispersion statistics (Section 2), and the statistics of distribution shape (Section 3). Within each section, the descriptive statistics, the estimates of their standard error, as well as the estimates of the confidence intervals are defined. As we point out, most of these estimates assume a normally distributed population. In these cases, using standard error estimates from samples taken from non-normal populations is detrimental to the inference made about the population's parameters. We return to this issue in Appendix B. For convenience, Table 1 (last page of the article) summarizes the relevant equations and

information given in the article regarding the descriptive statistics, standard error and confidence intervals.

Statistics of Central Tendency

Statistics of central tendency are estimates that aim to describe the population with a single value that represents the typical "central" score. Widely used measures of central tendency include the Pythagorean means: the arithmetic mean (often referred to as simply the mean), the harmonic mean, and the geometric mean. Other measures include the median and the mode. The mode will not be presented in this review as it has no agreed-upon estimate (see Robertson & Cryer, 1974) and no known standard error.

Arithmetic Mean

The arithmetic mean, or mean, represents the center of gravity of the dataset such that the data is balanced on both sides of the mean (Watier, Lamontagne & Chartier, 2010). The mean is affected by outliers and is therefore not a robust statistic. See Daszykowski, Kaczmarek, Heyden & Walczak (2007), Daszykowski, Kaczmarek, Stanimirova, Vander Heyden & Walczak (2007) and Walczak & Massart (1998) for a discussion on robust statistics. The mean of a dataset, noted $\bar{\mathbf{X}}$, is calculated using the following equation:

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \quad (1)$$

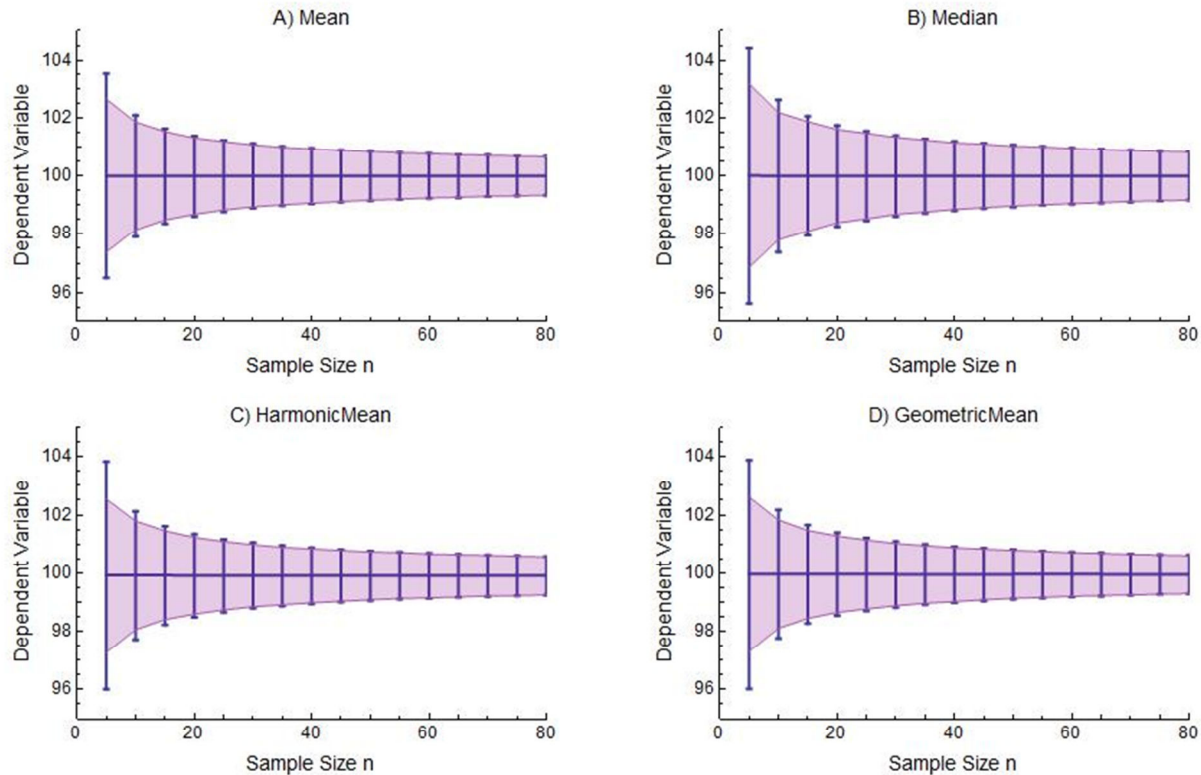
where \mathbf{X}_i represents the i^{th} value of the sample and n represents the sample size. The standard error of the mean ($\sigma_{\bar{\mathbf{X}}}$) is given by the following equation (Ahn & Fessler, 2003; Kendall & Stuart, 1977):

$$\sigma_{\bar{\mathbf{X}}} = \frac{\sigma_{\mathbf{X}}}{\sqrt{n}} \quad (2)$$

where $\sigma_{\bar{\mathbf{X}}}$ is the standard error of the mean and $\sigma_{\mathbf{X}}$ represents the population's standard deviation. In practical applications, as the population's standard deviation is often unknown, an estimated standard deviation is used, as shown in Equation (3):

$$SE_M = \hat{\sigma}_{\bar{\mathbf{X}}} = \frac{s_{\mathbf{X}}}{\sqrt{n}} \quad (3)$$

Figure 1 ■ Mean estimated central tendency (horizontal blue line) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Each point is based on 10,000 data points sampled from a normal distribution with a true mean of 100 and a true standard deviation of 3. Panel A) Mean; Panel B) Median; Panel C) Harmonic mean; Panel D) Geometric mean.



where s_x is the sample standard deviation and SE_M is the estimated standard error of the mean. The standard error of the mean is not based on the assumption of a normally distributed population (Kendall & Stuart, 1977) and is therefore applicable whatever the shape of the population's distribution. To obtain a confidence interval about the mean, a critical value must be taken from Student's t distribution, based on a confidence level, α . As the confidence interval is two tailed, the α value is divided by two. For this article, all simulations used a confidence interval of 95% so that $\alpha = 0.05$. The confidence interval equation for the SE_M is given by Equation (4) (Ahn & Fessler, 2003; Kendall & Stuart, 1977):

$$CI_{1-\alpha} = \bar{X} + SE_M \times t_{n-1}(\alpha/2, 1 - \alpha/2) \quad (4)$$

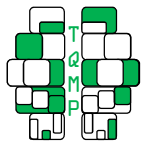
where $t_{n-1}(\alpha/2, 1 - \alpha/2)$ is based on $n - 1$ degrees of freedom. It returns both left ($\alpha/2$) and right ($1 - \alpha/2$)

critical values. As the t values are symmetrical about zero, we could equivalently write $CI_{1-\alpha} = \bar{X} \pm SE_M \times t_{n-1}(1 - \alpha/2)$. However, this is not true of all the CIs presented next. We will therefore use the more accurate notation of Equation (4).

Figure 1A presents the simulated and estimated standard error of the mean for increasing sample sizes. As seen, the results show that the confidence intervals obtained from an estimated standard error are very reliable estimates of the actual standard error. For smaller sample sizes (below $n = 10$), the 95% confidence interval for Equation (3) slightly overestimates the actual confidence interval.

Median

The median is defined as the point of a distribution where the number of data on each side of the median is equal. It is a robust alternative to the mean as it is little influenced by outliers. The value of the median is represented by \tilde{X} . The median can be found using the following equation (Kendall & Stuart, 1977):



$$\tilde{\mathbf{X}} = \begin{cases} \mathbf{X}_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{1}{2}(\mathbf{X}_{(\frac{n}{2})} + \mathbf{X}_{(\frac{n+1}{2})}) & \text{if } n \text{ is even} \end{cases} \quad (5)$$

where n is the sample size, $\mathbf{X}_{(i)}$ is i^{th} datum when the sample is sorted in increasing order. The standard error of the median (SE_{MD}) is estimated using the following equation:

$$SE_{MD} = \hat{\sigma}_{\tilde{\mathbf{X}}} = \sqrt{\frac{\pi}{2}} \frac{s_{\mathbf{X}}}{\sqrt{n}} \quad (6)$$

where $s_{\mathbf{X}}$ is the sample's standard deviation. As seen when comparing (6) with (3) the, SE_{MD} is $\sqrt{\pi/2} \approx 1.253$ larger than the SE_M . In other words, the median tends to be less reliably estimated than the mean (by 25%). This is one of the reasons that the mean is usually preferred over the median.

Unlike the mean, the standard error of the median assumes a normal distribution. To calculate confidence intervals, critical values must be taken from Student's t distribution. The confidence interval is given by the following equation:

$$CI_{1-\alpha} = \tilde{\mathbf{X}} + SE_{MD} \times t_{n-1}(\alpha/2, 1 - \alpha/2). \quad (7)$$

Figure 1B presents the simulated and estimated standard error of the median for increasing sample sizes. As seen, the results show that the estimated standard error using Equation (6) and (7) are very reliable estimates of the actual standard error. For smaller sample sizes (below $n = 15$), the 95% confidence interval from Equation (7) slightly overestimates the actual confidence interval.

Harmonic Mean

The harmonic mean is the reciprocal of the sum of reciprocals divided by n . With this descriptive statistic, effects of larger numbers are reduced as the reciprocal is inversely proportional to the datum's size. Conversely, numbers below 1 make for a large reciprocal due to the fact that numbers tending towards 0 have reciprocals tending towards infinity. A limitation to the harmonic mean, however, is that the data must be entirely positive. The harmonic mean is generally recommended to average rates. The harmonic mean, $\mathbf{H}_{\mathbf{X}}$, is calculated

using Equation (8) (Limbrunner, Vogel, Brown, 2000; Lam, Hung & Perrier, 1985):

$$\mathbf{H}_{\mathbf{X}} = \frac{n}{\sum_{i=1}^n \frac{1}{\mathbf{X}_i}} \quad (8)$$

The standard error of the harmonic mean SE_H is given by the following equation (Norris, 1940):

$$SE_H = \hat{\sigma}_{\mathbf{H}_{\mathbf{X}}} = \mathbf{H}_{\mathbf{X}}^2 \times \frac{s_{1/\mathbf{X}}}{\sqrt{n-1}} \quad (9)$$

where $s_{1/\mathbf{X}}$ is the standard deviation of the sample's reciprocals. The confidence intervals are given by Equation (10). Again, the critical value must be taken from the Student t table (Norris, 1940):

$$CI_{1-\alpha} = \mathbf{H}_{\mathbf{X}} + SE_{H_{\mathbf{X}}} \times t_{n-1}(\alpha/2, 1 - \alpha/2) \quad (10)$$

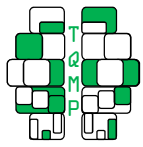
Figure 1C presents the simulated and estimated standard error of the harmonic mean for increasing sample sizes. Results show a similar tendency as the mean where smaller n lead to overestimated width of the actual CI for the harmonic mean. Sample sizes that are larger than $n = 10$ seem to lead to reliable estimates.

Geometric Mean

The geometric mean is another statistic of central tendency. It is represented by the n^{th} root (as opposed to the quotient by n) of the product (rather than the sum) of its parts. It is useful when the range of measures spans many magnitudes. Just like the harmonic mean, values in the dataset must be positive for the geometric mean to be calculated. The geometric mean, noted by $\mathbf{G}_{\mathbf{X}}$ is given by the following equation (Norris, 1940; Alf & Grossberg, 1979):

$$\mathbf{G}_{\mathbf{X}} = \sqrt[n]{\prod_{i=1}^n \mathbf{X}_i} \quad (11)$$

The standard error of the geometric mean (SE_G) is given by the following equation (Norris, 1940; Alf & Grossberg, 1979):



$$SE_G = \hat{\sigma}_{G_X} = G_X \times \frac{s_{\log(X)}}{\sqrt{n-1}} \quad (12)$$

where $s_{\log(X)}$ is the standard deviation of the logarithm of the data. The confidence interval is calculated using Equation (13), where again, the critical value must be taken from the t table:

$$CI_{1-\alpha} = G_X + SE_G \times t_{n-1}(\alpha/2, 1 - \alpha/2) \quad (13)$$

Figure 1D presents very similar results as the arithmetic mean and the harmonic mean where smaller sample sizes lead to overestimated estimates. For n larger than 10, Equation (13) leads to reliable estimates.

Statistics of Dispersion

The statistics of central tendency quantify the approximate central location of the distribution but do not describe the spread of the data. Unlike those indicators, statistics of data dispersion are used to describe the extent to which values are located around the center of the distribution. In this section we will assess the standard errors for variance, the standard deviation, the median absolute deviation and the interquartile range, with a small detour by the quantile.

Variance

The variance is a measure of dispersion that describes the extent to which values of a dataset are located around the mean. A large variance means that the dataset is widely dispersed across its values whereas a lower variance indicates that the dataset is spread tightly around the mean. Variance of a sample, noted as s_X^2 , is estimated with the following equation:

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (14)$$

The standard error of variance, SE_{S^2} , is given by (Ahn & Fessler, 2003; Kendall & Stuart, 1977):

$$SE_{S^2} = \hat{\sigma}_{s_X^2} = \sqrt{\frac{2}{n-1}} s_X^2 \quad (15)$$

where s_X^2 is the variance of a given sample and n is the sample size. Note that this estimator is derived from the assumption of a normally distributed population.

The sampling distribution of the variance is asymmetrical; the right tail tends to be longer than the left tail, owing to the squaring. For that reason, critical values cannot be taken from a t distribution as the distribution is not symmetrical about zero. To calculate confidence intervals, the critical values are taken from a χ^2 distribution at a specified α level. The confidence intervals are computed using a ratio with the following equation:

$$CI_{1-\alpha} = s_X^2 \frac{n-1}{\chi_{n-1}^2(\alpha/2, 1 - \alpha/2)} \quad (16)$$

in which $\chi_{n-1}^2(\alpha/2, 1 - \alpha/2)$ is read on a χ^2 table with $n-1$ degrees of freedom. The mean value of a χ^2 distribution with ν degrees of freedom is ν so that $\frac{n-1}{\chi_{n-1}^2}$ returns numbers around 1. For example, a 95% confidence interval for a sample of 10 subjects would have an $\alpha = 0.05$, $n = 10$, and $\chi_{10-1}^2(0.05/2, 1 - 0.05/2) = \{2.7, 19.023\}$ so that $(n-1)/\chi_9^2(0.05/2, 1 - 0.05/2) = \{3.333, 0.473\}$ (i.e. the lower bound is about 50% of the observed variance and the upper bound is 3.3 times larger than the observed variance).

Figure 2A presents the simulated and estimated standard error of the variance for increasing sample sizes. Results show that estimates found with Equation (15) are unreliable for sample sizes smaller than $n = 20$ where the equation highly overestimates the upper bound of the CI and slightly underestimates the lower bound. Therefore, Equation (16) should not be used for sample sizes smaller than 20.

Standard Deviation

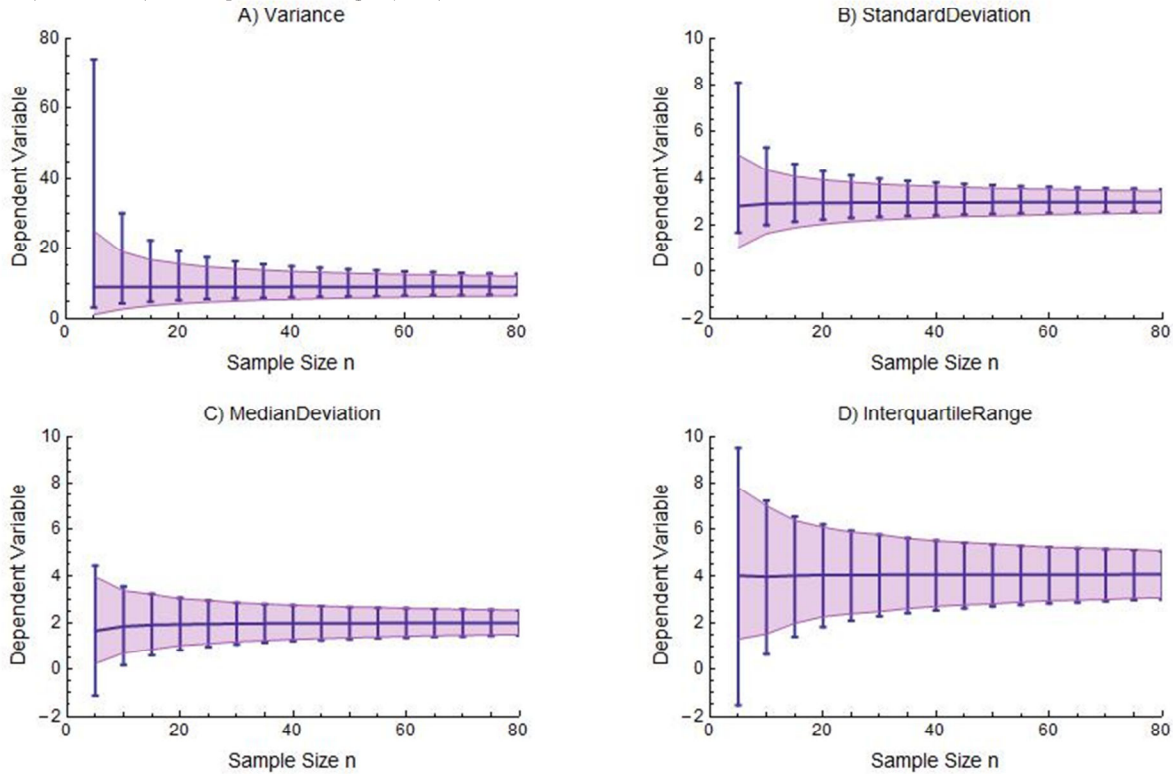
The standard deviation is calculated by taking the squared root of the variance. The standard deviation represents the typical deviation from the mean expressed in the same measurement unit as the data. The standard deviation, noted s_X , is calculated using:

$$s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (17)$$

The standard error of standard deviation (SE_{SD}) is given by:

variance. Just like its squared counterpart, the above

Figure 2 ■ Mean estimated dispersions (horizontal blue line) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Details are the same as in Figure 1. Panel A) Variance; Panel B) Standard deviation; Panel C) Median Absolute Deviation (MAD); Panel D) Interquartile range (IQR).



$$SE_S = \hat{\sigma}_{s_x} = \frac{\sqrt{2}}{\sqrt{n-1}} s_x \sqrt{\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}} \quad (18)$$

where $\Gamma(n) = (n-1)!$ for integer values of n . However,

$$\sqrt{\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}}$$

converges rapidly towards $\frac{1}{2}$. From this convergence, we get Equation (19) as an appropriate estimator.

$$SE_S = \hat{\sigma}_{s_x} \approx \frac{1}{2} \frac{\sqrt{2}}{\sqrt{n-1}} s_x = \frac{s_x}{\sqrt{2(n-1)}} \quad (19)$$

Note that even if the standard deviation is the square root of the variance, the standard error of the standard deviation is not the square root of the standard error of

estimator is derived by assuming a normally distributed population. To calculate confidence intervals, the critical values must be taken from the χ distribution (not to be confused with the χ^2 distribution) at the specified α value. Statistical tables of the χ distribution are uncommon but most statistical packages can compute quantiles based on this distribution. The confidence interval is given by:

$$CI_{1-\alpha} = s_x \frac{\sqrt{n-1}}{\chi_{n-1}(\alpha/2, 1-\alpha/2)} \quad (20)$$

Figure 2B presents the simulated and estimated standard error of the standard deviation for increasing sample sizes. Similar to the variance, Equation (20) is unreliable for sample sizes smaller than $n = 20$. To compare with the previous example, $\sqrt{n-1}/\chi_9(0.025, 97.5) = \{1.8256, 0.6878\}$.

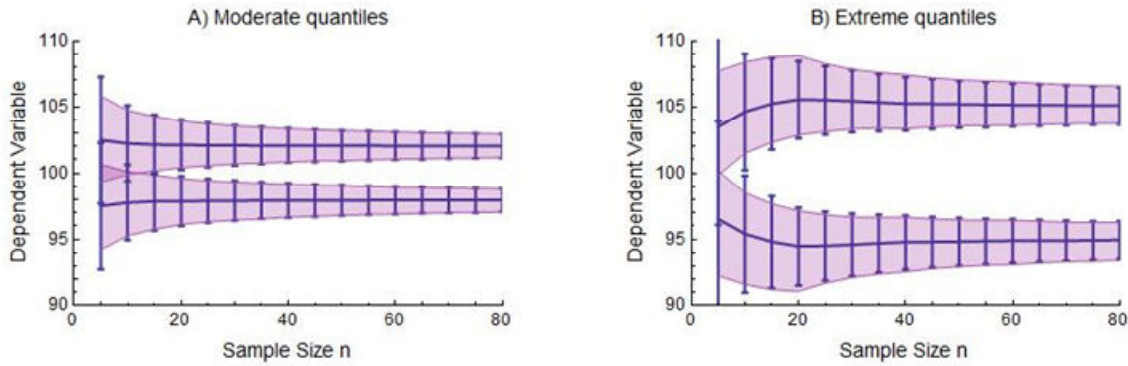


Figure 3 ■ Mean estimated quantiles (horizontal blue lines) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Details are the same as in Figure 1. Panel A) Moderate quantiles (25% and 75%); Panel B) Extreme quantiles (5% and 95%).

Median Absolute Deviation

Median absolute deviation is to the median what the standard deviation is to the mean. It is the median of deviations between the raw data and the median. It represents to what extent the data deviates from the median irrespective of the direction of the deviation (Gia & Hung, 2001):

$$\text{MAD}_{\mathbf{X}} = \text{median}(|\mathbf{X}_i - \tilde{\mathbf{X}}|) \quad (21)$$

where $\text{MAD}_{\mathbf{X}}$ is the median absolute deviation. For a normal population with μ and $\sigma_{\mathbf{X}}$ parameters, the population's $\text{MAD}_{\mathbf{X}}$ is given by $\frac{\sigma_{\mathbf{X}}}{z(0.75)}$ in which $z(0.75)$ is the 75th percentile of a standardized normal distribution ($z(0.75) \approx 0.6745$). The standard error of the median absolute deviation SE_{MAD} is given by (Gia & Hung, 2001):

$$SE_{\text{MAD}} = \hat{\sigma}_{\text{MAD}_{\mathbf{X}}} = \sqrt{\frac{2}{\pi}} \frac{s_{\mathbf{X}}}{\sqrt{n}} \quad (22)$$

where $s_{\mathbf{X}}$ is the standard deviation of a given sample and n is the sample size.

To calculate confidence intervals, the critical values are taken from Student's t distribution. The confidence interval is given by:

$$\text{CI}_{1-\alpha} = \text{MAD}_{\mathbf{X}} \pm SE_{\text{MAD}} \times t_{n-1}(\alpha/2, 1 - \alpha/2) \quad (23)$$

Figure 2C presents the simulated and estimated standard error of the median absolute deviation for increasing sample sizes. Note that the population $\text{MAD}_{\mathbf{X}}$ is $3/0.6745 = 2.02$. Contrary to variance and standard deviation, the estimates of $\text{MAD}_{\mathbf{X}}$ are rapidly reliable: results show that only for sample sizes smaller than or equal to $n = 10$, the estimated standard error are to unreliable and overestimated.

Quantiles

Before presenting the last statistic of dispersion, the interquartile range, we will present the quantile as this measure is instrumental in the computation of interquartile ranges. Quantiles are divisions of a dataset into equal-sized groups. For example, quantile 0.50 is the median, quantile 0.25 is the first quartile and quantile 0.01 is the first percentile (percentiles are frequently expressed as percentage). In a normal distribution, the 95% percentile is 1.645 and the 97.5% percentile is 1.960, which are commonly used as critical z values.

Various methods exist to estimate a quantile for a given percentile p . One convenient method when working with a potentially small sample is to interpolate between two entries. If we define r as the fractional part of p times $(n - 1)$, the operators $\lceil x \rceil$ and $\lfloor x \rfloor$ as the floor and ceiling of the quotient x , then $\mathbf{Q}_{\mathbf{X}}(p)$ is defined as:

$$\mathbf{Q}_{\mathbf{X}}(p) = \begin{cases} \mathbf{X}_{(p(n+1))} & \text{if } p(n+1) \text{ is an integer} \\ \mathbf{X}_{(\lfloor p(n+1) \rfloor)} + r(\mathbf{X}_{(\lceil p(n+1) \rceil)} - \mathbf{X}_{(\lfloor p(n+1) \rfloor)}) & \text{otherwise} \end{cases} \quad (24)$$

where $0 < p < 1$ is the desired quantile (Cousineau,

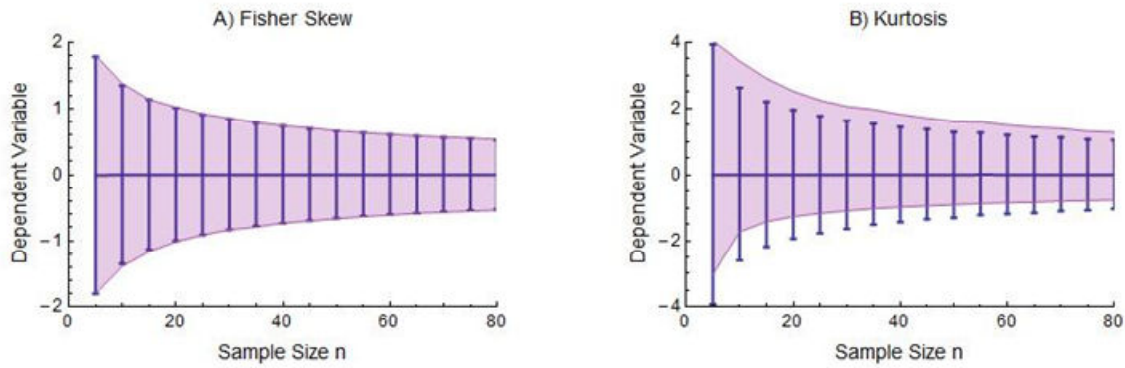


Figure 4 ■ Mean estimated shape of a distribution (horizontal blue line) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Details are same as in Figure 1. Panel A) Fisher skew; Panel B) Kurtosis.

2009).

The standard error of quantiles SE_Q is given by (Jin, Fu & Xiong, 2003; Kendall & Stuart, 1977):

$$SE_Q = \hat{\sigma}_{Q_X(p)} = \frac{s_X \sqrt{p(1-p)}}{\sqrt{n} \times \phi(z(p))} \quad (25)$$

where $z(p)$ is the p^{th} quantile of the standard normal distribution (typically read on a z table) and $\phi(z(p))$ is the density function of the standard normal distribution evaluated at the position $z(p)$. For example, with $p = 0.5$, we find the standard error of the median, Equation (6), as $\phi(0) = 1/\sqrt{2\pi}$. The confidence interval of a quantile is given by:

$$CI_{1-\alpha} = \frac{1}{\sqrt{n}} \times SE_Q \times t_{n-1}(\alpha/2, 1 - \alpha/2) \quad (26)$$

Because the present formula is dependent on both n and p , rather than on only n , simulations were performed with moderate (0.25 and 0.75) and extreme (0.05 and 0.95) values of p . Figures 3A and 3B present the observed and estimated CI of the quantile for $p = 0.25/0.75$ and $p = 0.05/0.95$ respectively. The results show that as p becomes extreme, the estimations are less accurate because of a sparseness of observations. Therefore, the CI for a quantile is accurate for $n > 10$ when p is moderate ($p = 0.25$) and for $n > 25$ when p is extreme ($p = 0.05$).

Interquartile Range

Quartiles are given by dividing the dataset into 4

segments. The interquartile range (IQR) is calculated by subtracting the 1st quartile from the 3rd quartile. The result is the range of data that covers the central 50% of the dataset. The value of the IQR is given by:

$$IQR_X = Q_X(0.75) - Q_X(0.25) \quad (27)$$

The standard error of the IQR is the standard error of the difference between two quartiles and is given by (Kendall & Stuart, 1977):

$$SE_{IQR} = \hat{\sigma}_{IQR_X} = \frac{s_X}{2\sqrt{n}\phi(z(0.25))} \quad (27)$$

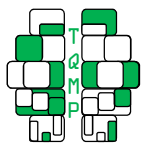
(note that $\phi(z(0.25)) \approx 0.31777$) The confidence interval for the IQR is therefore given by:

$$CI_{1-\alpha} = IQR_X + SE_{IQR} \times t_{n-1}(\alpha/2, 1 - \alpha/2) \quad (28)$$

Figure 2D presents the simulated and estimated values of IQR for varying sample sizes. The results show that Equation (28) is fairly accurate as soon as n exceeds 20. For smaller samples, the estimated standard error overestimates the true variability of the interquartile range.

Statistics of Distribution Shape

There are two measures that are commonly used to describe the shape of a distribution: Fisher skewness



and kurtosis. Note that the statistics of dispersion and central tendency can remain unchanged even in conditions where the shape is not normal since those indicators are independent from the shape of a distribution.

Skewness

Skewness is used to measure the asymmetry of a distribution. A distribution with a large (away from zero) coefficient of skewness is one in which most of its data is distributed on one side of the mean. The larger the coefficient, the more data there is on one side of the mean. A negatively skewed distribution signifies that the distribution has a long tail leading to the right with most of the data on the left side of the mean. The Fisher skew is estimated using (Doane & Steward, 2011; Joreskog, 1999; Joanes & Gill, 1998):

$$g_{\mathbf{X}} = \frac{\sqrt{n(n-1)}}{n-2} \frac{\frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^3}{s_{\mathbf{X}}^3} \quad (29)$$

The standard error for skewness (SE_S) can be estimated with (Kendall & Stuart, 1977):

$$SE_{SK} = \hat{\sigma}_{\hat{\gamma}_{\mathbf{X}}} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (30)$$

where n is the sample size. The standard error for skewness is estimated solely from the sample size and does not require the standard deviation of the sample. To calculate confidence intervals, the critical values are taken from the normal distribution at the α specified using:

$$CI_{1-\alpha} = g_{\mathbf{X}} + SE_{SK} \times z(\alpha/2, 1 - \alpha/2) \quad (31)$$

Figure 4A presents the simulated standard error of skewness for increasing sample sizes. The results show that unlike all statistics shown in this article, smaller sample sizes do not lead to unreliable estimates. The estimates found with Equation (31) are extremely reliable in every condition tested.

Kurtosis

Kurtosis is another measure of the shape of a distribution. Kurtosis is a measure of how "fat" the tails

are, or how rapidly their probability vanishes towards zero. It is restricted to be larger than -2; 0 is the kurtosis of a normal distribution. A low value of kurtosis signifies that the tails vanish very quickly (Ballanda & MacGillivray, 1988) whereas a high value of kurtosis signifies that the tails seem flatter, therefore vanishing at a slower pace. The kurtosis, noted as $b_{\mathbf{X}}$, is calculated using the following equation (Doane & Steward, 2011; Joreskog, 1999; DeCarlo, 1997; Joanes & Gill, 1998):

$$b_{\mathbf{X}} = \frac{(n+1)}{(n-2)(n-3)} \left((n+1) \left(\frac{\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^4}{n s_{\mathbf{X}}^4} - 3 \right) + 6 \right) \quad (32)$$

The standard error for kurtosis (SE_K) is estimated using:

$$SE_K = \hat{\sigma}_{b_{\mathbf{X}}} = 2 \times SE_{SK} \times \sqrt{\frac{n^2 - 1}{(n-3)(n+5)}} \quad (33)$$

where again, n represents the sample size and SE_{SK} is the standard error of skewness calculated in (30). A common way to calculate confidence intervals for kurtosis utilizes Equation (34). The critical values are taken from the normal distribution at a given α value.

$$CI_{1-\alpha} = b_{\mathbf{X}} + SE_K \times z(\alpha/2, 1 - \alpha/2) \quad (34)$$

Figure 4B, presents the simulated standard error of kurtosis for increasing sample sizes. The results show that the confidence intervals shown in Equation (34) exaggerate the possibility of a large kurtosis but underestimate the risk of a negative kurtosis. However, contrary to all the other estimators seen in this review, the errors of estimation do not tend to disappear with larger sample size. The fact that Equation (34) is unreliable regardless of sample size is a concern because it can lead to erroneous inference. The problem comes from the use of the z critical values that are symmetrical when kurtosis is not. In Appendix A, we present an alternative to the z critical values.

General discussion

This article is a review of the standard errors associated

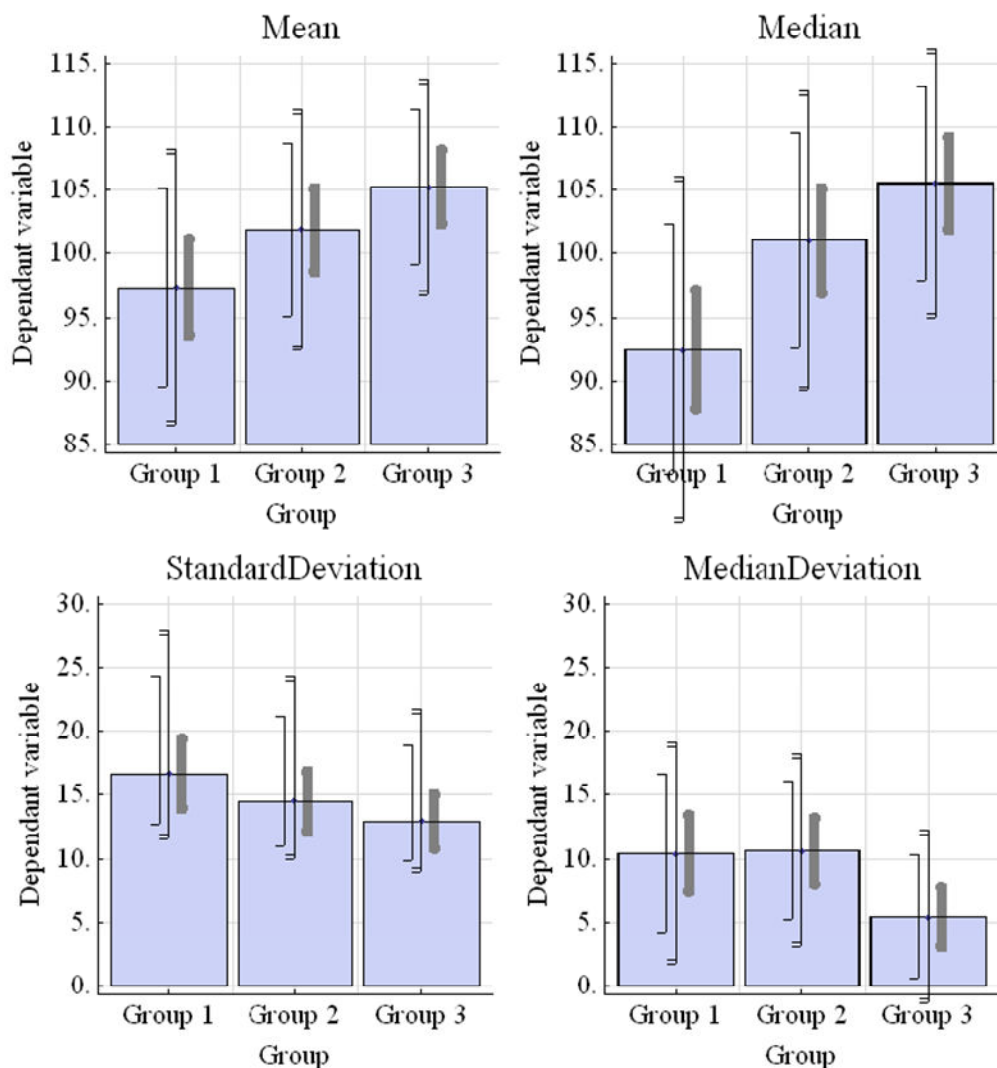


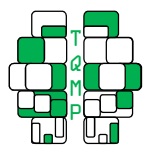
Figure 5 ■ Results of three groups in a fictitious task. Confidence intervals given are left: 95%, center (double-stroke): 99%, and right (thick): 1 standard error. Descriptive statistics of the groups are Panel A) Mean; Panel B) Median; Panel C) Standard Deviation; Panel D) Median Absolute Deviation.

with widely used descriptive statistics. It also provides an examination of the reliability of different equations that estimate the standard error. The results show that all measures evaluated are somewhat unreliable for small sample sizes, except for the standard error of skewness. An unusual result was found for the kurtosis standard error; to remedy this issue, the authors proposed a new approach to set critical values using a lognormal distribution (Appendix A).

All the estimates, other than the means, have standard error estimators that assume a normally distributed population. In Appendix B, we present a case study with a population varying on skew. This infringement of the normality assumption leads to

unreliable standard error for quantile estimates. Therefore, normality assumption should always be evaluated in the data before interpreting results. However, the impact of the infringement of this assumption had fewer effects as sample sizes increase in size, being, large than $n = 300$ at the least.

A point worth mentioning is that most of the standard error and confidence interval estimates tend to overestimate the range of plausible results when they were not perfectly accurate. Hence, these intervals are conservative and avoid wrongful impressions. This conservative bias is appropriate: seeing a difference where there might not be one would be unwarranted, more so in a discipline that publishes mostly rejections



of the null hypothesis (Yong, 2012).

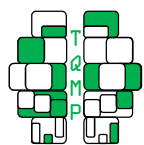
An application of standard errors and confidence intervals concern error bars around the measured statistics in summary plots. Confidence intervals can allow for inference on bar graphs: confidence intervals built with standard errors hints to the significance level of differences for example.. Confidence interval built with critical values (commonly used values are 95% and 99%) show a significant difference when the limit of each interval does not encompass the other group's measured statistic. Imagine a fictitious task where we measure the mean, the median, the standard deviation and the median absolute deviation for three independent groups. As seen in Figure 5, group 1 and 2 do not show any significant difference when comparing means or medians, yet group 1 and 3 are significantly different (standard error bars do not overlap nor do the 95% intervals encompass the other group's statistic). When comparing statistics of spread between groups we can infer that there is no significant difference as all the bars overlap each other between groups, supporting the homogeneity of variance assumption.. Standard errors or confidence intervals should always accompany a plot showing summary statistics. Loftus (1993, 1996), Beaulieu-Prévost (2006), Wilkinson and the APA Task Force on statistical inference have argued for its importance. Loftus (1993) went even further to argue that once reliable error bars are shown, null hypothesis testing is no longer necessary. Although this position is no longer considered strictly exact (Baguley, 2012, Volker and Loftus, 2012) error bars still offer considerable insights to the data. This review covered tools to improve the quality of data representation by facilitating the use of statistical indicators other than the mean. We hope this review will motivate researchers to gather more information from their data by using a wider array of indicators rather than the commonly used mean and standard deviation.

Authors' note

The first two authors have contributed equally in the making of this article.

References

- Ahn, S., & Fessler, J. (2003). Standard error of the mean, variance and standard deviation estimators (Report). Ann Arbor, MI, USA: EECS Department, University of Michigan.
- Alf, E. F., & Grossberg, J. M. (1979). The geometric mean: Confidence limits and significance tests. *Perception & Psychophysics*, 26(5), 419-421.
- Altman, D. G., & Bland, J. M. (2005). Standard deviation and standard errors. *Education and Debate*, 331, 903.
- Ballanda, K. P., & MacGillivray, H. P. (1988). Kurtosis: a Critical Review. *The American Statistician*, 42(2), 111-119.
- Bareiro, P. L., & Albandoz, P. J. (2001). Population and Sample. *Sampling Techniques* (Report). Spain: Management Mathematics for European Schools, University of Seville.
- Beaulieu-Prévost, D. (2007). Confidence Intervals: From tests of statistical significance to confidence intervals, range hypotheses and substantial effects. *Tutorials in Quantitative Methods for Psychology*, 2, 11-19.
- Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers*, 36, 742-756.
- Cousineau, D. (2009). *Panorama des statistiques en psychologie: Introduction aux méthodes quantitatives*. Bruxelles: De Boeck.
- Cousineau, D., Engmann, S., & Mestari, Z. (submitted). Evidence for parallel coactivation with double and triple redundant targets, p. 1-49.
- Daszykowski, M., Kaczmarek, K., Stanimirova, I., Vander Heyden, Y., & Walczak, B. (2007). Robust SIMCA-bounding influence of outliers. *Chemometrics and Intelligent Laboratory Systems*, 87, 95-103.
- Daszykowski M., Kaczmarek, K., Vander Heyden, Y., & Walczak, B. (2007). Robust statistics in data analysis: A review of basic concepts. *Chemometrics and Intelligent Laboratory Systems*, 85, 203-219.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307.
- Doane, D. P., & Seward, L. E. (2011). Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*, 19(2), 1-18.
- Jin, X., Fu, M. C., & Xiong, X. (2003). Probabilistic error bounds for simulation quantile estimators. *Management Science*, 14(2), 230-246.
- Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(1), 183-189.
- Joreskog, K. G. (1999). Formulas for skewness and kurtosis. Scientific Software International. Retrieved from <http://www.ssicentral.com/lisrel/techdocs/kurtosis.pdf>.
- Kendall, S. M., & Stuart, A. (1977). The advanced theory



- of statistics. (4th ed., Vol. 1). London, Great Britain: Charles Griffin & Company Limited.
- Lam, F. C., Hung, C. T., & Perrier, D. G. (1985). Estimation of variance for harmonic mean half-lives. *Journal of Pharmaceutical Sciences*, 74(2), 229-231.
- Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyse data. *Current Directions in Psychological Science*, 5, 161-171.
- Loftus, G. R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, & Computers*, 25, 250-256.
- Limbrunner, J. F., Vogel, R. M., & Brown, L. C. (2000). Estimation of harmonic mean of a lognormal variable. *Journal of Hydrologic Engineering*, 5, 59-66.
- Mouri, H. (2013). Log-normal distribution from a process that is not multiplicative but is additive. *Physical Review E*, 88, 42124-1 - 42124-7.
- Norris, N. (1940). The Standard Errors of the Geometric and Harmonic Means and Their Application to Index Numbers. *The Annals of Mathematical Statistics*, 11(4), 445-448.
- Olsen, C. H. (2003). Review of the use of statistics in Infection and Immunity. *Infection and Immunity*, 71(12), 6689-6692.
- Pham-Gia, T., & Hung, T. L. (2001). The Mean and Median Absolute Deviations. *Mathematical and Computer Modelling*, 34, 921-936.
- Robertson, T., & Cryer, J. D. (1974). An iterative procedure for estimating the mode. *Journal of the American Statistical Association*, 69(348), 1012-1016.
- Streiner, D. L. (1996). Maintaining standards: Differences between the standard deviation and standard error, and when to use each. *The Canadian Journal of Psychiatry*, 41(8), 498-502.
- Volker, H. F., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review*, 19, 395-404.
- Walczac, B., & Massart, D. L. (1998). Multiple outlier detection revisited. *Chemometrics and Intelligent Laboratory Systems*, 41, 1-15.
- Watier, N. N., Lamontagne, C., & Chartier, S. (2011). What does the mean mean? *Journal of Statistics Education*, 19(2), 1-20.
- Wilkinson, L., & the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Yong, E. (2012). Bad copy. *Nature*, 485, 298-300.

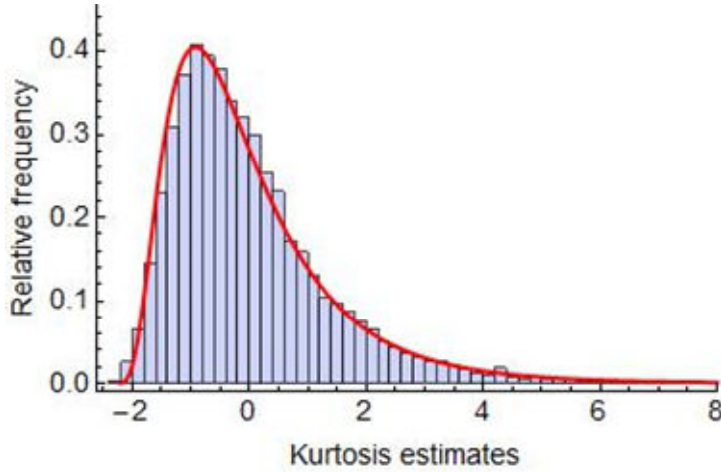
Appendix A: new estimator of the confidence intervals for kurtosis

The usual confidence interval estimator for kurtosis is inappropriate for all sample sizes; the estimator assumes that the distributions of estimates are symmetrical about the true kurtosis. However, as was seen in Figure 4 (the shaded area in Panel B), this is not the case. In this appendix, we present an alternative method to retrieve critical values that replaces the z values previously presented.

The new critical values are taken from a lognormal distribution. Mouri (2013) has shown that the sum for skewed deviates tend towards a lognormal distribution. This result conflicts with the central limit theorem which indicates that such a sum should tend towards a normal distribution. However, as Mouri (2013) showed, the sum of deviates tends faster towards a lognormal distribution than towards a normal distribution. This is the case regardless of the fact that as the number of addend increases, the normal distribution will ultimately be the asymptotic distribution.

Regarding kurtosis, we note that it is composed of a sum of n addend, each created by taking a random value raised to the power of four. Such a big exponent implies that the results will have a distribution with a very strong negative skew. To confirm this observation, we simulated 10,000 samples with $n = 10$ taken from a normal distribution, and recorded kurtosis for each. Figure A.1 shows the result of these simulations.

Figure A.1 ■ Distribution of 10,000 kurtosis estimates based on sample size of $n = 10$ taken from a normal distribution with a mean of 100 and a standard deviation of 3. The full red line is the proposed lognormal distribution with parameters: $\mu = \log(2)$, $\sigma = \frac{1}{2}SE_K$, and shift = \min_{b_K} .



The lognormal distribution has three parameters (Cousineau, Brown & Heathcote, 2004): the shift of the distribution, the μ parameter, and the σ parameter. The first parameter indicates how negative kurtosis can be, knowing that this measure is bounded from below. The theoretical distribution with the most negative kurtosis is the Bernoulli distribution with $\pi = \frac{1}{2}$ for which the kurtosis is -2. Due to the fact that we use a formula for kurtosis that is corrected for estimation bias (Eq. 32), the minimum depends on the sample size: it is -6 for a sample size of 4, -4 for a sample size of 5, etc. The minimum tends towards -2 for infinite sample size. In general, the formula (by replacing the ratio $((1/n) \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^4) / s_{\mathbf{X}}^4 - 3$ with -2 in Eq. 32) is given by:

$$\min_{b_K} = \frac{2(n-1)}{n-3} \quad (\text{A.1})$$

which depends only on the sample size. As for the μ parameter, results from Monte-Carlo simulations suggest that using 0.6940 returns good estimates. Finally, σ seems well characterized by $1/2 SE_K$, where SE_K is given by Equation (33). The lognormal distribution with such parameters can be simplified into a unit lognormal distribution due to the following properties where the quantile function of the lognormal distribution is written as $LN_{(\mu, \sigma)}(\alpha)$:

$$LN_{(\mu, \sigma)}(\alpha) = \exp(\mu)(LN_{(0,1)}(\alpha))^\sigma \quad (\text{A.2})$$

Also note that $\exp(0.6940)$ equals 2.0.

Thus, alternative 95% confidence intervals are found by looking for the 2.5th percentile and the 97.5th percentile of a lognormal distribution displaced by \min_{b_K} units to the left. Hence:

$$CI'_{1-\alpha} = b_K + 2 \times (LN_{(0,1)}(\alpha/2, 1-\alpha/2))^{(1/2)SE_K} - \min_{b_K} \quad (\text{A.3})$$

in which $LN_{(0,1)}$ denotes a standard lognormal distribution (with parameters μ equal to zero and σ equal to 1).

This new formulation for kurtosis confidence intervals was tested with simulations identical to the ones in the main text. The results are presented in Figure A.2. As seen, the 95% confidence intervals are very reliable past

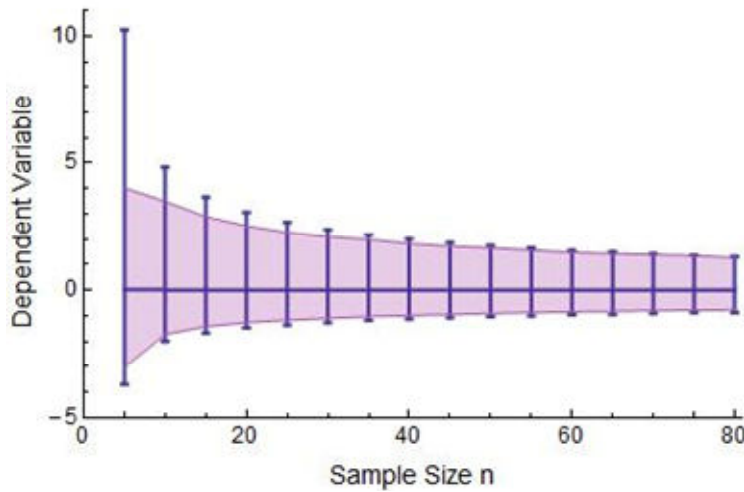


Figure A.2 ■ Mean estimated central tendency (blue line) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Details are the same as in Figure 1.

Table A.1 ■ Quantiles for the standard lognormal distribution $LN_{(0,1)}$.

Left Critical Values					Right Critical Values				
0.005	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99	0.995
0.0761	0.0977	0.1409	0.1930	0.2776	3.6022	5.1803	7.0991	10.2405	13.142

$n = 50$. In addition, the new confidence intervals are conservative as they never underestimate the real dispersion of the kurtosis estimates. Understandably, this new method of estimation is very bad for a sample size of $n = 5$: note that kurtosis requires a minimum of 4 data to be computed.

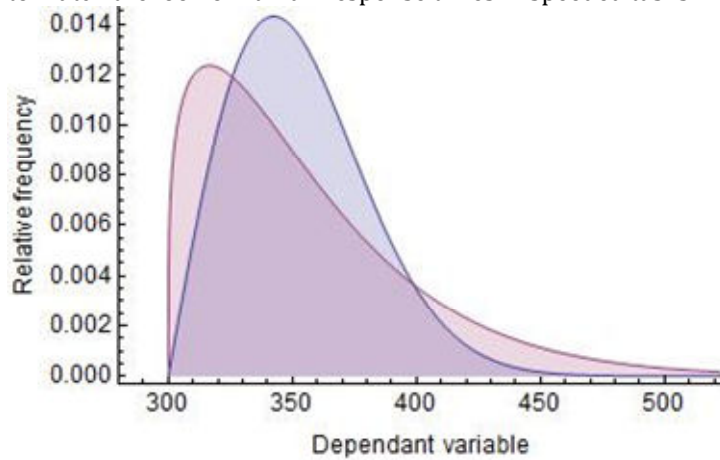
This new estimator requires a statistical table of the lognormal distribution. It is provided in Table A.1. This proposed method is considerably more reliable and should therefore replace the existing method that utilizes the z table.

Appendix B: Case example in which the normality of distribution assumption is violated

The present section examines the accuracy of the estimator for the SE_M (Eq. 4) and the SE_Q (Eq. 25) when the normality assumption is not respected. The simulations will show the error associated with an estimators that requires a normally distributed population, which is the case for standard errors of quantiles. The distribution used is a Weibull distribution with a scale parameter of $\beta = 60$, a shift parameter of $\alpha = 300$, and a shape parameter of either $\gamma = 2$ or $\gamma = 1.25$. The Weibull distribution was chosen since it is commonly postulated in cognitive science: response times in simple tasks such as the same-different task have distributions that are similar looking to the Weibull distribution. In addition, the left tail of a Weibull distribution has a weaker kurtosis than a normal distribution whereas the right tail has a stronger kurtosis. A review by Cousineau, Mestari, & Engmann (submitted) found an average shape parameter for the best-fitting Weibull distributions of 1.7 for tasks that required speeded responses. For illustrative purposes, we used two shape parameters: a shape of 1.25 to illustrate a strongly skewed distribution, and a shape parameter of 2.00 to illustrate a moderately skewed distribution. Figure B.1 presents the shape of the Weibull distribution for the chosen parameters.

Figure B.2 presents the simulated and estimated standard error of the mean for increasing sample sizes whereas Figure B.3 presents the simulated and estimated standard error of the quantile ($p = 0.25$ and $p = 0.75$) for increasing sample sizes. The results gathered on the mean (Figure B.2) show an accurate estimation of the actual standard error of the population mean. This is not surprising because Equation (4) does not postulate the normality of the distribution. The estimation therefore remains unaffected by the amount of skew in the population (Panel A vs. Panel B). Unlike standard errors of the mean, Figure B.3 shows that standard errors for the first and third

Figure B.1 ■ Two Weibull distributions: one with an important skew ($\gamma = 1.25$), the other with moderate skew ($\gamma = 2.00$). Both have a shift of 300 and a scale of 60. The parameters were chosen to match the look of human response times in speeded tasks.



quartiles are impacted by skew: estimations are unreliable with small samples (below $n = 20$) for moderately skewed populations. Conversely, when the skew of the distribution is large, estimates are never adequate: the estimators strongly exaggerate the standard error of the lower quartile and they strongly underestimate the standard error of the upper quartile. These results are apparent for even moderately large sample sizes ($n = 80$) and huge sample sizes are required before the observed dispersion of the estimates matches the dispersion suggested by the estimated standard error ($n > 300$).

(Continues on next page)

Figure B.2 ■ Mean estimated Mean (blue line) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Each point is based on 10,000 data points sampled from a Weibull distribution with a true mean of 353.2 and 355.9 for the panel A and B respectively. Panel A) Moderately skewed Weibull distribution; B) Strongly skewed Weibull distribution.

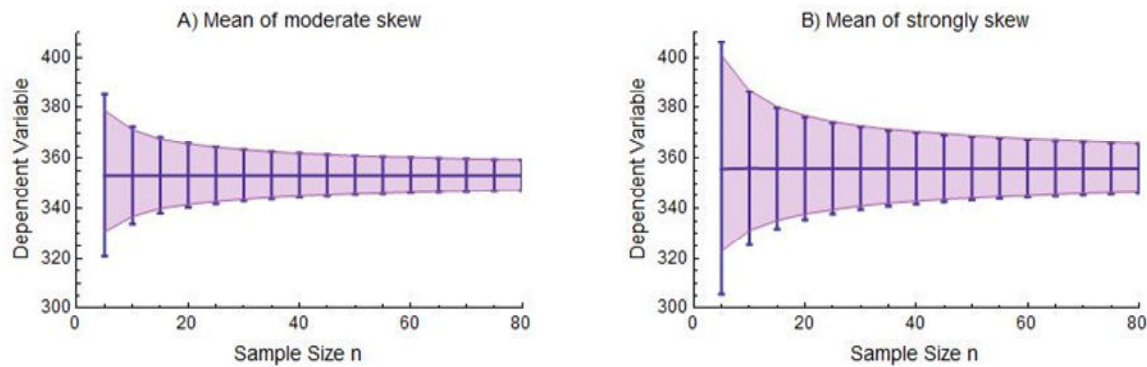
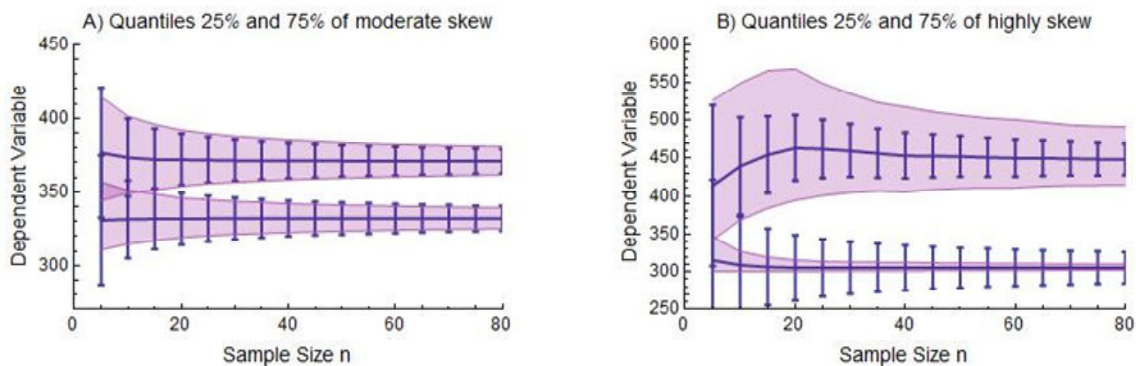


Figure B.3 ■ Mean estimated 25% and 75% quantiles (blue line) as well as estimated (error bars) and actual (shaded area) 95% confidence intervals as a function of sample size. Details are the same as Figure B.2.



Citation

Harding, B., Tremblay, C., & Cousineau, D. (2014). Standard errors: a review and evaluation of standard error estimators using Monte Carlo simulations. *The Quantitative Methods for Psychology*, 10 (2), 107-123.

Copyright © 2014 Harding, Tremblay & Cousineau. This is an open-access article distributed under the terms of the *Creative Commons Attribution License (CC BY)*. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 15/12/13

Table 1 follows...

Table 1 ■ Summary of equations, standard error estimator, and confidence intervals for descriptive statistics given in the article.

Descriptive Statistics	Equation	Standard Error	Confidence Interval
Statistics of Central Tendency			
Mean	$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$	$SE_M = \hat{\sigma}_{\bar{\mathbf{X}}} = \frac{s_{\bar{\mathbf{X}}}}{\sqrt{n}}$	$CI_{1-\alpha} = \bar{\mathbf{X}} + SE_M \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Median*	$\tilde{\mathbf{X}} = \begin{cases} \mathbf{X}_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{1}{2}(\mathbf{X}_{(\frac{n}{2})} + \mathbf{X}_{(\frac{n}{2}+1)}) & \text{if } n \text{ is even} \end{cases}$	$SE_{MD} = \hat{\sigma}_{\tilde{\mathbf{X}}} = \sqrt{\frac{2}{\pi}} \frac{s_{\tilde{\mathbf{X}}}}{\sqrt{n}}$	$CI_{1-\alpha} = \tilde{\mathbf{X}} + SE_{MD} \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Harmonic Mean	$\mathbf{H}_X = \frac{n}{\sum_{i=1}^n \frac{1}{\mathbf{X}_i}}$	$SE_H = \hat{\sigma}_{\mathbf{H}_X} = \mathbf{H}_X^2 \times \frac{s_{1/X}}{\sqrt{n-1}}$	$CI_{1-\alpha} = \mathbf{H}_X + SE_{HX} \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Geometric Mean	$\mathbf{G}_X = \sqrt[n]{\prod_{i=1}^n \mathbf{X}_i}$	$SE_G = \hat{\sigma}_{\mathbf{G}_X} = \mathbf{G}_X \times \frac{s_{\log(\mathbf{X})}}{\sqrt{n-1}}$	$CI_{1-\alpha} = \mathbf{G}_X + SE_G \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Statistics of Dispersion			
Variance*	$s_{\mathbf{X}}^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^2$	$SE_{S^2} = \hat{\sigma}_{s_{\mathbf{X}}^2} = \sqrt{\frac{2}{n-1}} s_{\mathbf{X}}^2$	$CI_{1-\alpha} = s_{\mathbf{X}}^2 \frac{n-1}{\chi_{n-1}^2(\alpha/2, 1 - \alpha/2)}$
Standard Deviation*	$s_{\mathbf{X}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^2}$	$SE_s = \hat{\sigma}_{s_{\mathbf{X}}} \approx \frac{s_{\mathbf{X}}}{\sqrt{2(n-1)}}$	$CI_{1-\alpha} = s_{\mathbf{X}} \frac{\sqrt{n-1}}{\chi_{n-1}(\alpha/2, 1 - \alpha/2)}$
Median Absolute Deviation*	$MAD_{\mathbf{X}} = \text{median}(\mathbf{X}_i - \tilde{\mathbf{X}})$	$SE_{MAD} = \hat{\sigma}_{MAD_{\mathbf{X}}} = \sqrt{\frac{2}{\pi}} \frac{s_{\mathbf{X}}}{\sqrt{n}}$	$CI_{1-\alpha} = MAD_{\mathbf{X}} \pm SE_{MAD} \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Quantile*	$\mathbf{Q}_X(p) = \begin{cases} \mathbf{X}_{(p(n+1))} & \text{if } p(n+1) \text{ is an integer} \\ \mathbf{X}_{(\lfloor p(n+1) \rfloor)} + r(\mathbf{X}_{(\lfloor p(n+1) \rfloor)} - \mathbf{X}_{(\lfloor p(n+1) \rfloor - 1)}) & \text{otherwise} \end{cases}$	$SE_Q = \hat{\sigma}_{\mathbf{Q}_X(p)} = \frac{s_{\mathbf{X}} \sqrt{p(1-p)}}{\sqrt{n \times \phi(z(p))}}$	$CI_{1-\alpha} = \frac{1}{\sqrt{n}} \times SE_Q \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Interquartile Range*	$IQR_{\mathbf{X}} = \mathbf{Q}_X(0.75) - \mathbf{Q}_X(0.25)$	$SE_{IQR} = \hat{\sigma}_{IQR_{\mathbf{X}}} = \frac{s_{\mathbf{X}}}{2\sqrt{n\phi(z(0.25))}}$	$CI_{1-\alpha} = IQR_{\mathbf{X}} + SE_{IQR} \times t_{n-1}(\alpha/2, 1 - \alpha/2)$
Statistics of Distribution Shape			
Skewness*	$\mathbf{g}_X = \frac{\sqrt{n(n-1)}}{n-2} \frac{\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^3}{s_{\mathbf{X}}^3}$	$SE_{SK} = \hat{\sigma}_{\mathbf{g}_X} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$	$CI_{1-\alpha} = \mathbf{g}_X + SE_{SK} \times z(\alpha/2, 1 - \alpha/2)$
Kurtosis*	$\mathbf{b}_X = \frac{(n+1)}{(n-2)(n-3)} \left((n+1) \left(\frac{\frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^4}{s_{\mathbf{X}}^4} - 3 \right) + 6 \right)$	$SE_K = \hat{\sigma}_{\mathbf{b}_X} = 2 \times SE_{SK} \times \sqrt{\frac{n^2-1}{(n-3)(n+5)}}$	$CI_{1-\alpha} = \mathbf{b}_X + SE_K \times z(\alpha/2, 1 - \alpha/2)$
Corrected Kurtosis*			$CI'_{1-\alpha} = \mathbf{b}_X + 2 \times (LN_{(0,1)}(\alpha/2, 1 - \alpha/2))^{(1/2)SE_K} - min_{\mathbf{b}_X}$

Note: Estimates that assume a normally distributed population are marked by an asterisk (*); \mathbf{X}_i denotes the i^{th} datum; $\mathbf{X}_{(i)}$ denotes the i^{th} datum after the data have been sorted in increasing order.